Primordial Spectral Distortions Originating from Cosmological Perturbations

A thesis submitted to The University of Manchester for the degree of
Doctor of Philosophy
in the Faculty of Science and Engineering

2023

Thomas Kite

Jodrell Bank Center of Astrophysics
School of Natural Sciences,
Department of Physics and Astronomy

Contents

Abstract 13

Declaration 14

Copyright Statement 15

Acknowledgements 16

Supporting Publications 17

1 From Zero to Cosmology 19

- 1.1 Geometry and Expansion 20
  - 1.1.1 Tensors and curvature 22
- 1.2 FLRW Geometry 26
  - 1.2.1 Relaxing the flatness assumption. 29
- 1.3 The Cosmic Inventory 29
  - 1.3.1 The radiation era 30
  - 1.3.2 The matter era 31
  - 1.3.3 The dark energy era 32
  - 1.3.4 Compact form for Friedmann's equation 33
- 1.4 Early Universe in ΛCDM 33

2 Shedding Light on the Photons 37

- 2.1 Photon interactions 38

2.1.1 Compton Scattering (CS) 39
2.1.2 Bremsstrahlung emission/absorption (BR) 41
2.1.3 Double Compton Scattering (DC) 42
2.2 Spectral shapes 43
2.2.1 Blackbody spectrum 43
2.2.2 Temperature shifts 45
2.2.3 y-Distortions 47
2.2.4 μ-Distortions 51
2.3 Primordial origins of SDs 54
2.3.1 The Temperature Era 54
2.3.2 The μ Era 55
2.3.3 The **y** Era 56
2.3.4 The residual era 56
2.4 Spatial spectral distortions 57

3 Cosmological Perturbation Theory 59
3.1 Introduction 59
3.2 The Boltzmann Equation 61
3.2.1 The Liouville Operator 62
3.2.2 The collision operator 63
3.2.3 The Photon hierarchy 64
3.3 Einstein’s field equations 66
3.3.1 SVT Decomposition 67
3.3.2 Gauge Choice 70
3.3.3 Stress Energy Tensor 73
3.3.4 Bottom line 74
3.4 Evolution of primordial perturbations 75
3.4.1 Horizon crossing 75
3.4.2 Tight coupling 77

3.4.3 Diffusion scale 78
3.4.4 Free-streaming 80
3.5 Integral Equation Approach 81
3.6 The Primordial Power Spectrum 82

Preface to Published Work **85**

4 Paper I: Bridging the gap: spectral distortions meet gravitational waves **87**

5 Paper II: Clarifying transfer function approximations for the large-scale gravitational wave background in ΛCDM **89**

6 Paper III: Spectro-spatial evolution of the CMB I: discretisation of the thermalisation Green's function **91**

7 Paper IV: Spectro-spatial evolution of the CMB II: generalised Boltzmann hierarchy **93**

8 Paper V: Spectro-spatial evolution of the CMB III: transfer functions, power spectra and Fisher forecasts **95**

9 Conclusion **97**

A Mathematical tools **101**
A.1 Legendre transform 101
A.2 Spherical Harmonics 102
A.3 Fourier transform 102

Word count: 66675

List of Figures

2.1 A figure showing the COBE/FIRAS data for the average photon spectrum seen across the sky. The error bars are multiplied by a factor of 400, emphasising just how precise the spectral shape is. Data courtesy of NASA. 45

2.2 A plot showing how temperature shifts added to a blackbody simply shift the blackbody in predictable ways. e.g. the peak of the distribution shifts to higher frequencies and amplitudes with higher temperature. Dashed lines show the result of evaluating the blackbody distribution at the new temperature, revealing small innacuracies for 10% variations of temperature, which would be remedied by adding y-distortions (defined below). 47

2.3 A plot showing the relative changes in photon spectrum captured in the temperature shift spectrum $\mathcal{G}$. Here a negative value indicates a relative lack of photons, and a positive branch shows an excess. The amplitude of variations here are comparable to that seen in real CMB data. The dashed lines again show an exact evaluation, showing here that temperature shifts are sufficient to model these small variations. . 48

2.4 A figure showing how the superposition of two blackbody spectra does not give another blackbody, rather there will inevitably be some SD. In the top panel the blue and red lines show the blackbody spectra at two different temperatures, and the dotted line shows the blackbody at the average temperature between them. In the bottom panels the resulting mixed spectrum is shown, demonstrating that it follows the dotted line till the Wien tail, where a distortion is present. Figure adapted from Chluba et al. (2015b). 51

2.5 A plot visualising the three main spectral distortions as relative spectra. This means the negative branches show relative absence of photons while the positive branches are excess photons. Roughly speaking y distortions push photons from the blackbody peak into the high frequency tail, while μ distortions push photons from the low frequency tail towards the blackbody peak. 52

2.6 A plot showing the three era picture with energy branching ratios (Chluba, 2016). Deep into each era the corresponding branching ratio tends to unity, indicating the production of a single spectral shape from energy injection at that time. 55

3.1 A diagram adapted from Dodelson (2003) showing different components in the early Universe with relevant interactions. Dotted lines indicate photons being produced or absorbed in the reaction. All lines connecting to the metric are understood as gravitational interactions. . 61

3.2 A figure showing how scalar and tensor perturbations differ in their dynamics. Θ_0 shows a constant amplitude envelope within which is oscillates due to a standing wave behaviour with electrons. The ending of that wave is characterised by a growing quadrupole moment, which is a useful quantity for understanding Silk damping and blackbody mixing (Chluba et al., 2012; Silk, 1968). The tensor perturbation on the the other hand decays immediately after horizon crossing but at a slower rate. This causes a stretching of the relevant tensor window for SD constraints compared to scalar perturbations (Chluba et al, 2019; Chluba et al., 2015a). 75

List of Tables

1.1 A table displaying the main steps in deriving the equations governing an FLRW Universe. 27

Abstract

THOMAS KITE
DOCTOR OF PHILOSOPHY
APRIL 2023

The anisotropies in the light arriving from the Cosmic Microwave Background (CMB) have provided cosmologists with a wealth of information over the previous two decades. Arguably this dataset has allowed cosmology to evolve into the role of precision science that it occupies today. This is due to the theoretical, mathematical and computational tools that link these anisotropies to the primordial perturbations that seeded structure in our Universe.

Spectral distortions of the CMB present a separate and complementary source of information from the night sky. Consisting of small deviations from the otherwise blackbody shape of the CMB spectrum, these signals probe the thermal history of the Universe at earlier times than recombination, offering a insight into the physics of the primordial plasma.

This thesis will explore connections between primordial perturbations and spectral distortions from two main viewpoints: Firstly the sourcing of average distortions from the dissipation of energy arising from these perturbations, with a particular focus on tensorial perturbations (papers 1-2, chapters 4-5). Secondly the prospect of studying a spatially varying distortion signal across the CMB, analogously to the usual CMB anisotropies (papers 3-5, chapters 6-8).

In the aim of completeness chapter 1 will review the introductory cosmology necessary to interpret the other chapters. Chapter 2 will introduce the photon spectrum in detail, covering both temperature shifts and distortions to the blackbody spectrum irreconcilable with changes of temperature. Chapter 3 will introduce cosmological perturbation theory, the framework which explains how small perturbations evolve in the otherwise homogeneous expanding Universe.

Supervisor: Prof. Jens Chluba
Co-supervisor: Prof. Richard Battye
Advisor: Prof. Robert Beswick

Declaration

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright Statement

(i) The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the “Copyright”) and he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

(ii) Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made **only** in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

(iii) The ownership of any certain Copyright, patents, designs, trade marks and other intellectual property (the “Intellectual Property”) and any reproductions of copyright works in the thesis, for example graphs and tables (“Reproductions”), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproducitons.

(iv) Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=487), in any relevant Thesis restriction declarations deposited in the University Library, The University Library’s regulations
(see http://manchester.ac.uk/library/aboutus/regulations) and in The University’s policy on Presentation of Theses.

Acknowledgements

I have been fortunate to undertake this PhD under Jens' supervision. His never-ending enthusiasm towards our work served as a bright light at the end of every tunnel.

I also thank my collaborators, Andrea and Subodh, who always kept me on the right track and never lacked an encouraging word.

I am thankful for the rest of the "spectral distortion gang" at Manchester: Luke Hart, Lizzy Lee, Aditya Rotti, Boris Bolliet, Jiten Dhanda and Bryce Cyr. Our Covid Zoom calls, group dinners and conference trips made you feel like extended family.

The broader community at JBCA and the constant supply of coffee were always a source of inspiration. In particular I want to mention Francesco Pace for our many conversations about British weather, Dan Thomas for the great sense of clarity he always conveyed, Roke Cepeda-Arroita for his grand tales of mountaineering and spelunking, and Shruti Badole for our extensive conversations on Ed Sheeran.

I must also acknowledge the many friends that have supported me and taught me a lot over the years: Dan Holmes essentially taught me everything I know about coding. Ed Broadberry firstly taught me most of what I know about QFT and secondly how to dance, even if I don't know much about either. Harry Howell taught me how to (theoretically) fly a plane. Ryan Barouki taught me many country's flags, and to never stop seeing the fun in life or trying new things. Mack prepared me for the conference ski slopes, and tried teaching me how to deal with spiders. Finally Ben Symons, who let's say taught me a lot about the non-academic side of life.

I thank my mother Karen who sacrificed a lot for me to be in this position today, and who has always been supportive even if what I was talking about sounded like a foreign language half of the time.

Finally I have only humble and heartfelt gratitude for my partner Anna. Her kind words, encouragment and sympathetic joy for all my milestones have not only made the PhD experience *orders of magnitude* easier, but also far more enjoyable.

Supporting Publications

Bridging the gap: spectral distortions meet gravitational waves

Thomas Kite, Andrea Ravenni, Subodh Patil, Jens Chluba

2021 MNRAS, 505, p.4396-4405

Chapter 4 of this thesis.

The author performed the pre-tabulated calculation of window functions which allowed for the central calculations of the paper. Furthermore the various individual mechanisms for creating tensor perturbations and their constraints were evaluated by the author.

Clarifying transfer function approximations for the large-scale gravitational wave background in ΛCDM

Thomas Kite, Jens Chluba, Andrea Ravenni, Subodh Patil

2022 MNRAS, 509, p.1366-1376

Chapter 5 of this thesis.

The author led this project, including the development of numerical tools, and analysis of different physical regimes and effects on the tensor power spectra.

Spectro-spatial evolution of the CMB I: discretisation of the thermalisation Green's function

Jens Chluba, Thomas Kite, Andrea Ravenni

Submitted to JCAP

Chapter 6 of this thesis.

The author contributed to the development of the expanded spectral basis and relevant operator algebra. Furthermore he performed the numerical treatment and analysis of convergence against theory.

Spectro-spatial evolution of the CMB II: generalised Boltzmann hierarchy

Jens Chluba, Andrea Ravenni, Thomas Kite

Submitted to JCAP

Chapter 7 of this thesis.

The author contributed in studying the suitability of the spectral basis, in particular focusing on the convergence of various operators which become important at first order in perturbation theory. Furthermore the numerical results (presented in the following paper) provided important checks on this more theoretical work.

Spectro-spatial evolution of the CMB III: transfer functions, power spectra and Fisher forecasts

Thomas Kite, Andrea Ravenni, Jens Chluba

Submitted to JCAP

Chapter 8 of this thesis.

This paper was heavily focused on a numerical treatment and data visualisation of the previous companion paper's bottom line equations, all of which was led by the author.

1

From Zero to Cosmology

Cosmology is a rich and exciting branch of physics, offering a unique opportunity to study the Universe on the grandest scales. Through deceptively simple models we can make very tangible progress towards meaningful explanations of the structure of the Universe around us.

The undertaking of this subject, however, comes with many difficulties. On the one hand, it is a very diverse subject, demanding its followers to be multidisciplinary. Particle physics, statistical physics and GR are just a few of the many fields essential to modern cosmological theories. On the other hand there is an experimental difficulty, since unlike most branches of the physical sciences there is only one lab in which cosmology can be studied: our one Universe. Despite the empirical challenge, cosmologists have made enormous progress in gathering substantial and robust data sets. Thanks to this, recent decades have seen the field of cosmology develop into a precision science. A large part of this success is due to the wealth of data now gathered from the Cosmic Microwave Background (CMB) and the sophisticated mathematical and computational tools used to study it, both of which we will discuss in detail throughout this thesis. This effort in understanding the dynamics of the early Universe has led to the consolidation of a concordance model in cosmology, ΛCDM, which will be a large part of the focus in this introductory chapter.

1.1 Geometry and Expansion

> Let no one ignorant of geometry come under my roof
>
> –Plato

Studying the entirety of space and time is a grandiose task, and one which requires similarly grandiose tools. Our current best understanding of space-time is derived from General Relativity (GR), a complex yet elegant description of out Universe in almost entirely geometric terms. In this section we aspire to introducing the relevant equations and concepts of GR so we can understand perhaps the simplest yet most consequential fact in cosmology: the Universe is expanding.

We will introduce GR by alluding to its origin in differential geometry, however we will not use as much of the mathematical machinery demanded by such an abstract approach, instead concerning ourselves only with the bottom line tensorial equations. The former approach however is satisfying in that it emerges naturally from physical principles, and justifies why tensorial equations are so ubiquitous in Physics. For good introductions to the conceptual side of GR see the introductory chapters of Carroll (2004), for a quick introduction to the mathematics see Appendix A of Weinberg (2008), and for a full balanced introduction see Misner et al. (2017).

Special relativity fundamentally changed the way we saw our Universe, with moving clocks ticking at different rates and travellers disagreeing with the lengths of each others trains. Digging deeper, the true change is that we could no longer permit ourselves to have a simple space and time separation in our view of the world, but instead thinking of a united spacetime. This constitutes the first of the physical principles alluded to above.

The second key physical principle is that of coordinate invariance, in which we demand that results of any calculation we perform be independent of the choice of coordinate system, understanding that the latter choice is purely an *artefact* we project onto the world to describe it. For example, calculations performed in Cartesian coordinates (x,y,z) must give the same physical bottom line as in polar coordinates (r,θ,ϕ). A more nuanced example, itself related to special relativity, is the interchangability of electric and magnetic fields upon choosing a new coordinate frame in relative motion to the first.

The punch line of GR is that gravity itself can be understood as a purely coordinate effect. This was motivated by the equivalence principle, which highlights how an observer in free fall is indistinguishable from an observer at rest. We must therefore understand how two inertial observers can have relative accelerations towards one another, an apparent paradox that *simply* requires a curved manifold. Consider for example that two ants walking in a straight line on a flat 2D sheet will at most meet in one point, and their distance will only grow/shrink at a constant rate. Ants on a basket ball however can both walk in straight lines have have an accelerated distance function between them.

Mathematically speaking, if we take the two principles seriously then a natural model arises in which the Universe is, in an abstract sense, a four dimensional manifold ($\mathcal{M}$), which can **locally** be mapped with a coordinate system ($x : \mathcal{M} \to \mathbb{R}^4$), the components of which we recognise as space (x, y, z) and time (t), at least locally. Asking which physical phenomena can be described in this highly abstracted system (without using arbitrary coordinates) reveals something very beautiful: we can describe paths across the manifold, and those paths induce a concept of gradient. These become the tangent spaces – the vector spaces in which our tensorial equations live. The reason therefore that tensorial equations are so pervasive in physics is that they're the coordinate invariant way of describing the way things move across manifolds (which with the fourth dimension also means time evolution). A key connection in this statement is that the coordinates chosen on the manifold $\{x^\mu\}$ also induce a choice of tangent space basis $\{\partial_\mu \equiv \partial/\partial x^\mu \ \equiv \mathbf{e}_\mu\}$ (we further explain this notation in Sec. 1.1.1).

This coordinate invariance is then crucially important in GR where gravity is expressed through a non-trivial curved manifold. If we imagine a flat 2D surface ($x - y$ plane) then it is clear we can simply express our differential equations in a trivial cartesian frame $\{\partial_x, \partial_y\}$. The choice of non-trivial coordinates (e.g. polar $\{\partial_r, \partial_\theta\}$) is optional, and is understood to incur some non-trivial jacobian contributions to the relevant equations. In GR we are not guaranteed to have such a simple cartesian choice. We can, at most, choose a trivial Minkowski metric at a single point of the manifold. Complicated coordinates with corresponding jacobians are inevitable. Tensorial equations, in essence, ensure that our equations are correct no matter the coordinates chosen.

This idea is used heavily in Weinberg (2008), stated as follows:

(...) it is only necessary to write the equations in a form which is generally covariant – that is, whose form is independent of the spacetime coordinates used – and which reduce to the correct equations in the absence of gravitation. Such equations will be true in the presence of a gravitational field, because general covariance guarantees that they are true in any set of coordinates if they are true in any other set of coordinates, and the Equivalence Principle tells us that there is a set of coordinates in which the equations are true – the set of coordinates that is locally inertial at the spacetime location of the system in question.

– Steven Weinberg. Cosmology, Oxford University Press, 2008.

1.1.1 Tensors and curvature

In this thesis we follow conventional notation for tensors where greek letters indicate time and space indices $(0, 1, 2, 3)$ while latin indices are purely spatial $(1, 2, 3)$, except in the simple 2D examples where we will not be so careful. We follow Einstein summation convention, where repeated indices are summed: $A_\mu B^\mu = \sum_{\mu=0}^{3} A_\mu B^\mu$.

Loosely speaking an inexperienced reader can/should think of A, A_μ and $A_{\mu\nu}$ as a scalar, a vector and a matrix respectively. Whenever we cast into these more concrete mathematical structures we will do so with an arrow ($\longrightarrow$) rather than an equivalence symbol to emphasise that it incurs a choice of basis, and thus a loss of generality.

In this coordinate independent tensorial language (otherwise simply called covariant notation) it is important to distinguish between covariant “down” indices (A_μ) and contravariant "up" indices (A^μ). In particular a well-formed tensor equation only contracts indices when one is covariant and the other contravariant. Readers should relate this to the idea that inner product of vectors require one row and one column vector, e.g.

$$\begin{bmatrix} a & b & c \end{bmatrix} \cdot \begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix}, \tag{1.1}$$

or perhaps a reader more familiar with quantum mechanics should think of "bra-ket" notation where inner products require a *sandwich* $\langle A|B\rangle$. The more general notion in mathematics is known as the dual space to a vector space, where dual vectors are explicitly defined to map vectors to real numbers ($v^* \in \mathbb{V}^*$ such that $v^* : \mathbb{V} \to \mathbb{R}$). The key here is that in the transpose example above we didn't modify the entries of the vector at all in the process. In general however there is a non-trivial modification in the mapping, which is contained in the metric $g_{\mu\nu}$:

$$g_{\mu\nu} A^\nu = A_\mu. \tag{1.2}$$

This layer of complication is necessary in describing manifolds with curvature, and even using non-Cartesian coordinates on flat manifolds (e.g. see upcoming polar coordinates example).

The metric $g_{\mu\nu}$ is, in more general terms, the tensor containing information about the manifold's structure. For example it defines the lengths across spacetime paths

$$\mathrm{d}s^2 \equiv g_{\mu\nu} \mathrm{d}x^\mu \mathrm{d}x^\nu. \tag{1.3}$$

Let's consider as an example the 2D plane, which if we express in the usual Cartesian (x, y) coordinates we have

$$g_{\mu\nu} \longrightarrow \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \tag{1.4}$$

which gives path lengths measured with

$$S = \int \sqrt{\mathrm{d}s^2} = \int \sqrt{\mathrm{d}x^2 + \mathrm{d}y^2}, \tag{1.5}$$

as expected. To emphasise the coordinate invariance here we will illustrate this example in polar coordinates (r, θ) where instead we have

$$g_{\mu\nu} \longrightarrow \begin{bmatrix} 1 & 0 \\ 0 & r^2 \end{bmatrix}, \tag{1.6}$$

and thus a path length

$$S = \int \sqrt{\mathrm{d}s^2} = \int \sqrt{\mathrm{d}r^2 + r^2\mathrm{d}\theta^2}. \tag{1.7}$$

Ultimately both of these expressions must give the same number regardless of coordinates – a simple example of "well-formed" tensor equations giving coordinate invariant properties. Note that in this example we only had a single coordinate θ stretching as a function of the other r. In principle we can have mixed components $g_{\mu\nu} \neq 0$ for $\mu \neq \nu$, but the metric must be symmetric $g_{\mu\nu} = g_{\nu\mu}$. Similar to this polar coordinates example, we will see in Sect. 1.2 that our Universe also has some components (space) stretching in terms of others (time). A second important point is that for the flat 2D plane we could start from a Cartesian system, and move to an *optional* non-trivial one. Consider for a moment the surface of the Earth (assuming this as perfectly spherical), which is also a 2D space, but does not allow for a simple Cartesian coordinate set.

Continuing the above example of a flat 2D sheet in the two different coordinate frames, we can introduce an extremely important concept of the covariant derivative Heuristically one can think of this as the well-formed tensor equivalent of simple differentiation in a direction. This derivative is simply ∂_μ in the Cartesian coordinates, but takes a new form ∇_μ in the coordinate independent language. Consider the polar coordinates example, where if you walk in the direction of basis vector $\mathbf{e}_\theta$, your local direction of $\mathbf{e}_\theta$ and $\mathbf{e}_r$ changes. Walking in the $\mathbf{e}_r$ direction you do not get any angle changes, but the $\mathbf{e}_\theta$ direction has a relative stretching. These changes are encoded in the Christoffel symbols $\Gamma^\lambda_{\mu\nu}$, which can be read as "the change of the λ component of basis vector $\mathbf{e}_\mu$ upon walking in direction of basis vector $\mathbf{e}_\nu$". These coefficients are formally given by

$$\Gamma^\lambda_{\mu\nu} = \frac{1}{2} g^{\lambda\rho} \left(\partial_\mu g_{\nu\rho} + \partial_\nu g_{\rho\mu} - \partial_\rho g_{\mu\nu} \right). \tag{1.8}$$

This backs up our previous statement that the metric encodes the manifold's structure, since these coefficients only depend on the metric and its derivatives. Two things are noteworthy here: firstly the symbols are symmetric in lower indices $\Gamma^\lambda_{\mu\nu} = \Gamma^\lambda_{\nu\mu}$, and secondly we are being careful in not referring to these objects as tensors, since they are not well formed tensor equations (they depend on partial derivatives and not covariant derivative – we are in the process of defining the latter!).

As an example, in the polar coordinates case we find

$$\Gamma^r_{\theta\theta} = -r, \qquad \Gamma^\theta_{r\theta} = \Gamma^\theta_{\theta r} = \frac{1}{r}. \tag{1.9}$$

Reading these carefully, the first states that walking in direction $\mathbf{e}_\theta$ changes $\mathbf{e}_\theta$ in the $\mathbf{e}_r$ direction by $-r$. Additionally the second term says that walking in that same direction changes $\mathbf{e}_\theta$ in the $\mathbf{e}_\theta$ direction by $1/r$. These two statements together amount to the angular change we expect to see, e.g. walking a quarter circle from $\theta = 0$ to $\theta = \pi/2$ in this plane makes the $\mathbf{e}_\theta$ vector transition from pointing up ($\mathbf{e}_y$) to pointing left ($\mathbf{e}_x$).

With changes of vector direction in place, the covariant derivative can be written as

$$\nabla_\lambda T^\mu_\nu = \partial_\lambda T^\mu_\nu + \Gamma^\mu_{\lambda\rho} T^\rho_\nu - \Gamma^\rho_{\lambda\nu} T^\mu_\rho, \tag{1.10}$$

where in general we extend this definition to as many positive terms as we have contravariant indices, and as many negative terms as covariant indices.

With all the formalism in place, we can introduce the tensorial equation at the heart of cosmology is Einstein's field equation, which is built up from the metric. In natural units[1] it is given by

$$G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = 8\pi G\, T_{\mu\nu}, \tag{1.11}$$

where $T_{\mu\nu}$ is the stress-energy content of the Universe, and the Ricci scalar R is the contracted Ricci tensor $R_{\mu\nu}$

$$R_{\mu\nu} = \frac{\partial \Gamma^{\lambda}_{\mu\lambda}}{\partial x^{\nu}} + \frac{\partial \Gamma^{\lambda}_{\mu\nu}}{\partial x^{\lambda}} + \Gamma^{\kappa}_{\mu\lambda}\Gamma^{\lambda}_{\nu\kappa} - \Gamma^{\kappa}_{\mu\nu}\Gamma^{\lambda}_{\lambda\kappa}. \tag{1.12}$$

The Ricci tensor is itself a contraction of the Riemann tensor, which encodes information about the curvature of the manifold. Importantly, it is only built from $g_{\mu\nu}$ and its derivatives, hence the statement that the metric contains the necessary information about the structure of the manifold. To remedy this awfully brief introduction we will further explore the geometry ($G_{\mu\nu}$) in sect. 1.2 and the cosmic inventory ($T_{\mu\nu}$) in sect. 1.3.

1.2 FLRW Geometry

In order to effectively use the machinery introduced in the previous section we must first drastically reduce the complexity of the general field equations Eq.(1.11). Other fields in physics teach us to use symmetries whenever confronted with a daunting problem like this, and in cosmology we have the Copernican principle to help us choose these symmetries. The principle essentially states that the Universe should not contain *special places* on large scales. Another way of stating this is to say that the Universe should be completely homogeneous and isotropic on average. Further to these symmetries, we add one more simplification which is experimentally verified: the Universe appears to be flat. We note that by this we mean spatial flatness. If one imposed the generalisation of flatness to 4 dimensional spacetime then there would be no gravitational effects, and cosmology would be rather boring. Later, in Sect. 1.2.1 we provide the metric without the assumption of flatness, and explain how this impacts the bottom line results in a simple way. These simplifications leave us with one option for the universal geometry, governed by the Friedmann–Lemaître–Robertson–Walker (FLRW) metric. Letting the spatial 3-metric have the shorthand $\mathrm{d}\boldsymbol{x}^2 = \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2$ we can write the metric in the following way:

---

[1] Unless otherwise stated we will use natural units throughout this thesis

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + a(t)^2\mathrm{d}\boldsymbol{x}^2 = a(t)^2(-\mathrm{d}\eta^2 + \mathrm{d}\boldsymbol{x}^2), \tag{1.13}$$

where $a(t)$ is the scale factor, a parameter accounting for the expansion of the Universe by allowing spatial distances to grow or shrink in time (a is defined to be $= 1$ at current time). In the second line we introduced conformal time η, defined by $a\mathrm{d}\eta = \mathrm{d}t$, corresponding to a coordinate frame where the time dimension dilates together with spatial expansion. The other common measure of time used in this report is redshift z defined by $1 + z = 1/a$, which is defined in relation to redshifted light from distant galaxies.

Taking the metric corresponding to this line element and propagating it through Einstein's equations is a simple process, the key points of which are summarised in table 1.1.

Christoffel Symbols	$\Gamma^0_{ij} = a\dot{a}\delta_{ij}$ $\Gamma^i_{0j} = \Gamma^i_{j0} = \frac{\dot{a}}{a}\delta_{ij}$ All others = 0
Ricci Tensor	$R_{00} = -3\frac{\ddot{a}}{a}$ $R_{0i} = R_{i0} = 0$ $R_{ij} = (2\dot{a}^2 + a\ddot{a})\delta_{ij}$
Ricci Scalar	$R = 6(\frac{\ddot{a}}{a} + \frac{\dot{a}^2}{a^2})$
Einstein Tensor	$G_{00} = 3\frac{\dot{a}^2}{a^2}$ $G_{0i} = G_{i0} = 0$ $G_{ij} = -(\dot{a}^2 + 2a\ddot{a})\delta_{ij}$

Table 1.1: A table displaying the main steps in deriving the equations governing an FLRW Universe.

The Christoffel symbols in this metric show that there is only spacetime curvature associated with the time dimension, since $\Gamma^i_{jk} = 0$.[2] Furthermore, if we had chosen to derive these in conformal time then we would find $\Gamma^0_{00} = \frac{a'}{a} \neq 0$, showing that conformal time dilates/contracts as it advances forward.[3]

What about the stress energy tensor? The form compatible with this geometry and coordinate choice has $T_{00} = \rho$ and $T_{ij} = p\delta_{ij}$, where we have anticipated some of the discussion in Sect. 1.3 by identifying these with energy density and pressure respectively. A fluid allowing a description in terms of just these two variables are known as perfect fluid, and are demanded by the symmetries of the Copernican principle.

By equating the results of table 1.1 and the stress-energy tensor, as per Einstein's equation $G_{\mu\nu} = 8\pi G T_{\mu\nu}$, we get the Friedmann equations:

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho, \tag{1.14}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3p), \tag{1.15}$$

which govern how the Universe will expand subject to its stress-energy content. The second equation is particularly important for modern cosmology, since it gives information on the acceleration of the Universe's expansion. In particular we have the scenarios

$$\frac{\ddot{a}}{a} \begin{cases} > 0 \ \text{(accelerated growth)} & \text{if} \quad p < -\rho/3, \\ = 0 \ \text{(stable expansion)} & \text{if} \quad p = -\rho/3, \\ < 0 \ \text{(accelerated collapse)} & \text{if} \quad p > -\rho/3. \end{cases} \tag{1.16}$$

While a fluid with negative pressure may seem odd from a classical point of view, we currently observe the Universe to be undergoing accelerated expansion, demanding such a fluid to exist (if our best models and theories are correct).

---

[2] Note that $\Gamma^i_{jk} = 0 \implies$ no curvature, but the contrary is not true, no curvature $\not\Longrightarrow \Gamma^i_{jk} = 0$. i.e. consider using polar coordinates in a flat space.

[3] A keen eyed reader will realise that statements around Christoffel symbols are evidently coordinate dependent, and thus they are not in fact *proper* tensors. All relevant physical quantities are built of combinations of Christoffel symbols such that the coordinate dependence cancel. These symbols are the only such imposters in this chapter – all other 'indexed' quantities can be trusted as coordinate independent tensors.

Either through a combination of Eqs. (1.14), (1.15), or by using the conservation law $\nabla_\mu T^\mu_\nu = 0$, one can also find

$$\dot{\rho} = -3\frac{\dot{a}}{a}(\rho + p), \tag{1.17}$$

which dictates how energy densities dilute or otherwise in the expanding Universe. We will use this equation in combination with the other Friedmann equations to model some limiting cases of universal expansions in Sect. 1.3.

1.2.1 Relaxing the flatness assumption.

As previously mentioned FLRW metrics are compatible with spatial curvature. The inclusion of this term in the metric gives

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + a(t)^2\Big(\frac{\mathrm{d}r^2}{1-\kappa r^2} + r^2\mathrm{d}\theta^2 + r^2\sin^2(\theta)\mathrm{d}\phi^2\Big), \tag{1.18}$$

where we have used spherical polar spatial coordinates. Here the term $\kappa = -1, 0, +1$ indicates hyperbolic/open, flat and spherical/closed geometry respectively. Propagating this through Einstein's equations would only yield a single additional factor, which can be absorbed into the previous ones with $\rho \to \rho - \frac{\kappa}{a^2}$. This essentially takes the form of a new fluid with equation of state $w = -1/3$.

1.3 The Cosmic Inventory

The RHS of Eq. (1.11) is an expression of the energy content of the Universe, which will be explored in this section. Firstly we will distinguish between perfect and imperfect fluids. The former are defined as a medium where at every point there is a locally inertial frame moving with the fluid in which the fluid is isotropic (Weinberg, 2008).

These fluids admit pressure and density as their only free parameters. These fluids are compatible with the Copernican principle, and thus makes up the cosmic inventory on large (averaged) scales. The alternative is having imperfect fluids, which can contain bulk motions, shear stresses, etc. Fluids of this kind will occupy much of our attention in later chapters where we apply perturbation theory to the primordial plasma.

Perfect fluids are usually categorised with an equation of state relating pressure to density, $p = p(\rho) = w\rho$. Assuming that w is constant, Eq. (1.14) and Eq. (1.17) can be easily integrated to find

$$\rho(a) \propto a^{-3(1+w)}, \tag{1.19}$$

$$a(t) \propto t^{\frac{2}{3(1+w)}}, \tag{1.20}$$

where the proportionality constant would be set through some initial or boundary conditions. This simple derivation is surprisingly powerful, as we can now plug in w for the most common species in cosmological history. We will do so in a chronological order of importance, but to do so requires somewhat spoiling the punchline: different particle species dilute and diffuse at different rates, leading to a picture of the Universe where we have distinct transitions between eras where a single fluid dominated.

1.3.1 The radiation era

In the earliest moments of the Universe massless particles dominated the total cosmic energy budget. This means photons, but also includes neutrinos, which at these temperatures could be considered massless. Due to this we will sometimes refer more correctly to a phase of *relativistic* energy domination (see e.g. chapter 5, where such distinction is of key importance).

These massless particles typically have an equation of state $w = 1/3$, which when plugged into Eq. 1.19 and Eq. 1.20 yield

$$\rho(a) = \rho_0 a^{-4}, \tag{1.21}$$

$$a(t) \propto \sqrt{t}. \tag{1.22}$$

The first result can be understood as a combination of the volume changing ($V \propto a^3$) and the relative energy loss from the gravitational redshifting of light ($E = h\nu \propto a^{-1}$). The second result serves as a reminder to the discussion around Eq. 1.16, providing an example of a universe that expands, but at a slower rate as time goes on.

1.3.2 The matter era

The relative energy densities of massless particles not only dilute in ever-larger volumes and also redshift to lower frequencies. This means at some point in the Universe's history they lose importance relative to massive particles. Matter is divided into baryonic (here meaning electrons, neutrons and protons, unlike other branches of physics) and cold dark matter (elusive matter which is apparent only from gravitational effects). Matter is modelled as not exerting any pressure on cosmological relevant scales. Plugging $w = 0$ into Eq. 1.19 and Eq. 1.20 gives

$$\rho(a) = \rho_0 a^{-3}, \tag{1.23}$$

$$a(t) \propto t^{2/3}. \tag{1.24}$$

A similar discussion holds here as before: energy densities change because of volume changes, and matter decelerates expansion as expected.

It is still an open question whether this particle (if indeed CDM *is* a particle) sector interacts weakly or not - thus far cosmological probes have only been sensitive to its gravitational influence. It is worth noting that its extra gravitational influence is required at many different scales: from rotation curves of galaxies, to cosmic web structure, all the way to the composition of the primordial plasma. This ubiquitous presence of CDM provides an argument for a new source of mass, and not simply some misunderstood gravitational physics (Clifton et al., 2012).

1.3.3 The dark energy era

It was mentioned above that the Universe currently appears to be accelerating in its expansion. None of the fluids so far have that effect. In the concordance model of cosmology the late time acceleration occurs due to a strange new fluid called dark energy. This can be thought of as a cosmological constant, or as vacuum energy. The former isn't so much an explanation of physical phenomena as much as a simple realisation that a term $\Lambda g_{\mu\nu}$ can be added to Eq. 1.11 while keeping all other relations and derivations the same (you can think of this as an integration constant). The latter interpretation has a bit more physical meaning to it, however estimating the size of this constant from fundamental physics has provided infamously bad results (Adler et al., 1995).

Ignoring matters of interpretation, this is usually modelled as a fluid with $w = -1$, yielding the odd results

$$\rho(a) = \rho_0, \tag{1.25}$$

$$a(t) \propto \exp(Ht), \tag{1.26}$$

where the second equation broke with the tradition of the previous two subsections, since Eq. 1.20 is derived implicitly assuming $w \neq -1$. We see the reason it is called a cosmological **constant**, and how it certainly accelerates expansion – it does so exponentially.

More generally we can think of some fluid with $p \approx -\rho$, allowing for a dynamic model. This fluid would dilute by some small amount, and the acceleration would not be exponential, but would be positive. This model is generally referred to as dark energy (DE), but models of this type are not often distinguishable from a simple cosmological constant with current cosmological data. We will assume $p = -\rho$ in this thesis.

It will perhaps not be surprising to the reader that many attempts have been made at replacing either CDM or Λ with some theory of modified gravity (MG), since they both seem to be anomalies purely constrained to the gravitational force (Clifton et al., 2012).

1.3.4 Compact form for Friedmann's equation

Combining what we have learned in the previous sections provides a very simple and elegant form of Friedmann's equations. Firstly we choose to measure the densities relative to a critical density $\rho_c = \frac{3H^2}{8\pi G}$, defining $\Omega_\mathrm{x} = \rho_\mathrm{x}/\rho_c$. This is useful, because in a flat universe we have $\Omega_\mathrm{total} = \sum \Omega_\mathrm{x} = 1$. This means each individual Ω_x simply gives a relative contribution to a total energy budget. We note that even in a cosmology that isn't flat we can package the rest of the energy contribution into some fictitious fluid $\Omega_\kappa \propto 1/a^2$ as in Sect. 1.2.1, however we disregard that here. The bottom line version of Eq. (1.14) with these simplifications is

$$H = \frac{\dot{a}}{a} = H_0 \sqrt{\frac{\Omega_m}{a^3} + \frac{\Omega_r}{a^4} + \Omega_\Lambda}, \tag{1.27}$$

with H_0 giving the measured value of the expansion rate as seen today (this convention applies for most variables with subscript 0).

1.4 Early Universe in ΛCDM

> In the beginning the Universe was created. This had made many people very angry and has been widely regarded as a bad move.
>
> –Douglas Adams, *The Restaurant at the End of the Universe*

From the discussion in the previous section we can start to infer what the very early Universe looks like in ΛCDM. Other than the cosmological constant all the energy densities scale with inverse powers of the scale factor, which itself grows with time. The simple extrapolation backwards is then that the Universe becomes a more energetic environment at earlier times, as may be sensible intuitively (imagine compressing every planet, star and galaxy into ever smaller volumes!). At some crucial energy density you find that even a typical photon has the requisite energy to ionise atoms, and the entire Universe is thus a plasma. This primordial plasma is the complex environment where this thesis' work takes place. Crucially photons have a very small mean free path in this plasma, and thus the Universe was essentially opaque.

There is a thin time slice separating a Universe so hot that atoms are ionised, and a Universe cold enough to allow photons to travel freely. The moment this happens is known as recombination, referring to how the electrons recombine with protons to make neutral matter. Shortly after that there is the moment of last scattering, where each photon undergoes its last interaction, after which it will wander the cosmos almost certainly unhindered. When we observe the night sky we see a sphere of these photons in all directions, and call it the Cosmic Microwave Background (CMB) (Penzias & Wilson, 1965).

In this leading order homogeneous picture CMB must be spherical[4], as it is simply dictated by the distance travelled at the speed of light since the last scattering time. Furthermore it is a sphere for every observer because of relativity, no matter the location or relative velocity. The surface of photons itself is known as the last scattering surface (not to be confused with LSS, the large scale structure), however in this thesis it will often be called the CMB curtain, since it serves to obscure anything which might otherwise be seen earlier in time.

Due to redshifting the CMB is seen at an extremely low temperature of $T = 2.725$K (Fixsen et al., 1996). Furthermore, it takes that temperature equally in all directions (up to small variations we discuss below). The apparent paradox of opposite celestial poles taking the same temperature, a potential violation of causality, is a known problem proposed to be solved via *inflation* (Guth, 1981).

The statements above on average temperatures and sphericity all hold on average, and to a very large degree. This can be seen as a consequence of the Copernican principle, which dictates there should be no preferred location in the Universe (isotropy follows from this principle). Potentially the most impactful dataset in modern cosmology however is comprised of temperature anisotropies: tiny variations of this temperature around the sky. We will discuss in the following chapters how we can apply a perturbation theory approach to the equations discussed here. This perturbative approach is well justified experimentally, since the observed fluctuations in the CMB are typically of order $\delta T/T \sim 10^{-5}$.

---

[4] Small under- and over-densities delaying the free-streaming of light can break this symmetry

2

Shedding Light on the Photons

One humbling fact about modern cosmology is that almost the entirety of our extragalactic information and data have come from a single source: observing light. Whether we are counting galaxies, observing gravitational lenses or witnessing CMB photons, they all come down to intercepting the otherwise free streaming photons. Only two notable exceptions exist to this. Firstly, there are infrequent neutrino detections[1]. To provide a sense of magnitude, the nearby supernova SN1987A yielded a mere 25 neutrino observations across three different Observatories (Arnett et al., 1989). Secondly, there are detections of gravitational waves (LIGO Scientific Collaboration & Virgo Collaboration, 2016; Abbott et al., 2019) which have recently opened the door to a long studied possibility of gravitational wave cosmology (Caprini & Figueroa, 2018), a field that promises to revolutionise our understanding of the Cosmos (see chapter 5).

The importance of information locked in photons then cannot be overstated. This chapter will explore the information content of CMB photons, and describe the various interactions photons undergo which gets locked into different variables. The main focus here will concretely be the information locked in frequency distributions, which is a key part of the study of Spectral Distortions (SDs). See Lucca et al. (2019) for a helpful review on the topic, Chluba (2014b) for the science case and Chluba (2016) for ΛCDM estimates.

---

[1] Not to be mistaken with indirect inference about neutrino properties with cosmology (Abazajian et al., 2015).

The punchline of this chapter is that while the photons from the primordial Universe constitute an almost exact blackbody spectrum, there must be some deviations known as spectral distortions. Simply stated, for a system to arrive at thermal equilibrium there must be interactions between particle species. These interactions become more rare as the universe expands and cools. On its own this would not lead to spectral distortions, but we additionally know that some minimal amount of energy injection into the primordial plasma occurs from the dissipation of acoustic modes (see Chapter 3). This race between energy injection processes and thermalisation processes leads to a three era picture of the early Universe, in which observing a given spectral distortion shape informs us in which era it was created.

Before discussing what spectral distortions are and how they form in the early universe, it will be helpful to enumerate the main interactions that photons undergo in the primordial plasma.

2.1 Photon interactions

In chapter 1 we closed with the idea that the earliest moments in known cosmic history are characterised by an entire Universe in a state of hot plasma. To recap, photons have enough energy to completely ionise any neutral atoms, and consequently they do not travel very far before scattering off an energetic electron. This is a complex environment, where many governing forces are at play against the backdrop of cosmic expansion. As energy densities slowly decrease, interactions become more rare.

In this environment we are going to study the various interactions photons undergo, with special attention to energy-changing and photon-producing interactions. These two types of interaction attempt to establish equilibrium between the participating particles. For example, a scattering event between photons and hot electrons will tend to scatter photons to higher frequencies. Those same hot electrons can produce soft photons via Bremsstrahlung and double Compton events with protons. These two processes establish equilibrium between photons and electrons in a time and frequency dependent way.

We will study these interactions in the form of collision terms $C[...]$ given by

$$\frac{\mathrm{d}f_\gamma(p)}{\mathrm{d}t} = C[f_\gamma, f_\mathrm{e}], \tag{2.1}$$

which we will come to know in the following chapter as the Boltzmann equation. The function $f_\gamma(p)$ on the LHS is the occupation function for phase space distribution function of photons. This is a momentum (or frequency) dependent distribution, since we want to understand any exchange of energy between electrons and photons. The function f_e is a similar distribution for electrons. Stated in words, the collision terms express how the photon gas changes (in a momentum dependent way), and is in general a functional of both the photon and electron phase space distributions at a given time.

2.1.1 Compton Scattering (CS)

This process describes the interaction of photons with electrons[2]:

$$e(p) + \gamma(k) \longleftrightarrow e(p') + \gamma(k'), \tag{2.2}$$

where p and k represent the four-momenta of particles, using primes for outgoing particles. Being an interaction of fundamental particles this process would be studied from first principles in the framework provided by Quantum Electrodynamics (QED):

$$\begin{aligned} C[f_\gamma(k')]\Big|_{\mathrm{CS}} = \frac{1}{2E_\gamma(k')} \int \frac{\mathrm{d}^3p}{(2\pi)^3 2E_\mathrm{e}(p)} \int \frac{\mathrm{d}^3p'}{(2\pi)^3 2E_\mathrm{e}(p')} \int \frac{\mathrm{d}^3k}{(2\pi)^3 2E_\gamma(k)} \\ \times |M|^2 (2\pi)^4 \delta^{(4)}(p + k - p' - k') \\ \times \Big[f_\mathrm{e}(p') f_\gamma(k')(1 + f_\gamma(k)) - f_\mathrm{e}(p) f_\gamma(k)(1 + f_\gamma(k')) \Big], \end{aligned} \tag{2.3}$$

---

[2] Photons will interact in a similar way with protons, but the terms involved are typically inversely proportional to the mass of the charged particle, making γe^- interactions ~ 2000× stronger than γH^+ interactions.

where E_e and E are energies of electrons and photons respectively, $|M|^2$ is the squared matrix element which will in general depend on the momenta involved, and f_e is the occupation function for electrons, similar to what we defined for photons above. The terms like $\frac{d^3p}{(2\pi)^3 2E_\mathrm{e}(p)}$ are Lorentz invariant integration measures over the space of momenta, and the Dirac delta $\delta^{(4)}$ ensures energy-momentum conservation. The final term is sometimes called the statistical factor, and essentially captures the physics of photons entering (leaving) the k state from (to) k'. In particular the terms $1 + f_\gamma$ account for induced scattering, which are nonlinear effects. An explicit form for $|M|^2$ can be found in Jauch & Rohrlich (1976).

Various approximations of $|M|^2$ can be used to distil the above equation in limiting regimes. One such limiting case is Thomson scattering which allows for changes of momentum direction, but not magnitude. Since here we are interested in energy exchange between photons and electrons we proceed one step further and introduce the Kompaneets equation (Kompaneets, 1957):

$$C[f_\gamma]\Big|_\mathrm{CS} \approx (c\sigma_\mathrm{T} n_\mathrm{e}) \frac{\theta_\mathrm{e}}{x^2}\frac{\partial}{\partial x} x^4 \left[\frac{\partial}{\partial x} f_\gamma + \frac{T_\gamma}{T_\mathrm{e}} f_\gamma (1 + f_\gamma)\right], \tag{2.4}$$

where T_i is the temperature[3] of particle species i, $\theta_\mathrm{e} = \frac{k_\mathrm{B} T_\mathrm{e}}{m_\mathrm{e} c^2}$, $x = \frac{h\nu}{k_\mathrm{B} T_\gamma}$, σ_T is the Thomson cross-section and n_e is the electron number density. Moving forward we will use $c\sigma_\mathrm{T} n_\mathrm{e} = \dot{\tau}$. The Kompaneets equation is derived assuming a non-relativistic electron gas at temperature T_e, and expands the matrix elements to second order in energy exchange (see discussion on moments below), thus including *Doppler broadening* terms.

Some physics can be seen in this collision term: Firstly, notice that for $T_\gamma = T_\mathrm{e}$ the term in brackets vanishes for a blackbody distribution (defined later in Eq. (2.10)), showing that the equilibrium solution is a blackbody only when photons and electrons are in equilibrium. Consequently, the different cooling rates between photons and electrons will cause deviations from a blackbody (see Sect. 2.2.3). Secondly the form $\frac{1}{x^2}\partial_x(...)$ immediately shows that this reaction conserves photon number, since $\int x^2 C[f_\gamma]\Big|_{\rm CS}$ corresponds to a vanishing surface term.

---

[3] This is slightly subtle. The Kompaneets equation can be applied to photon spectra that are not exactly blackbodies, and hence don't have a well defined temperature. However the photon temperature within x cancels with the explicit factor of T_γ, leaving only T_e as the temperature scale, since electrons can be more safely approximated as following a Maxwell-Boltzmann distribution in cosmological contexts. Despite this it is desirable to have a fiducial temperature for photons, and in practice we usually choose a temperature $T_\gamma \equiv T_z = T_0(1 + z)$ which due to the redshifting will cancel a contribution on the left hand side of Boltzmann's equation (see Eq. 3.3)

Detailed work on CS can be found in reference Sarkar et al. (2019), where they define the CS kernel, $P(x \to x')$, a probability distribution dictating the likelihood of transitioning from some frequency x to another x'. With this they find

$$C[f_\gamma(x)]\Big|_{\rm CS} = \int \left\{ \frac{x'^2}{x^2} P(x' \to x)\; f_\gamma(x')[1 + f_\gamma(x)] - P(x \to x')\; f_\gamma(x)[1 + f_\gamma(x')] \right\} {\rm d}x'. \tag{2.5}$$

By further defining a *moment* of a function $\xi(x)$ as $\langle \xi(x) \rangle = \int P(x \to x')\, \xi(x)\, {\rm d}x'$, it can be shown that ξ is fully defined by its powers $\langle \xi^N \rangle$. The Kompaneets equation corresponds to an expansion of Eq. (2.5) to second moments, $\langle (\frac{x'}{x})^2 \rangle$, with some approximate form of the kernel. This method provides a clear path forward where improving estimations of the kernel, and going to higher moments of $\langle \frac{x'}{x} \rangle$, will improve the accuracy of calculations. The code `CSpack` (Sarkar et al., 2019) contains efficient and accurate predictions of these moments, for easy implementation into other branches of cosmology.

2.1.2 Bremsstrahlung emission/absorption (BR)

Classically this process involves the emission/absorption of a photon from a charged particle undergoing acceleration[4], although we will focus on the case of emission for this section. In the early Universe this is electrons accelerating in the presence of some charged ion, predominantly protons (ionised hydrogen):

$$e(p) + H^+(h) \longleftrightarrow e(p') + H^+(h') + \gamma(k). \tag{2.6}$$

---

[4] Brems strahlung translates from German to braking radiation

This again could be expressed using an integral equivalent to Eq. (2.3), with a different matrix element. Typically a different approach has been employed in cosmology, where the classical and non-relativistic formula given by Kramer is modified by some function called the Gaunt factor, $g_{\rm ff}$:

$$C[f_\gamma]\Big|_{BR} \approx \frac{8\pi}{3}\frac{e^6 h^2}{m_{\rm e}(k_{\rm B}T_{\rm e})^3}\frac{Z^2 N_{\rm e} N_{\rm H}}{\sqrt{6\pi m_{\rm e} k_{\rm B} T_{\rm e}}}\frac{e^{-x_{\rm e}} g_{\rm ff}(T_{\rm e},\nu)}{x_{\rm e}^3}\Big(1 - f_\gamma[e^{x_{\rm e}} - 1]\Big), \tag{2.7}$$

where $x_{\rm e} = xT_\gamma/T_{\rm e}$. For recent work on the calculation of $g_{\rm ff}$ in different temperature regimes see Chluba et al. (2020b), where Chluba et al. bridge the gap between non-relativistic and ultra-relativistic calculations with the Elwert-Haug cross section (Elwert & Haug, 1969). These calculations are difficult to perform, but with a series of precomputed tables and analytic approximations they manage to cover many orders of magnitude of phase space at $\approx$1% precision for atomic number $Z\leq10$, and better than 0.1% precision for $Z\leq2$. Similarly to `CSpack`, these strong numerical capabilities are packaged in `BRpack`, allowing other cosmologists to easily implement accurate BR predictions in their work.

2.1.3 Double Compton Scattering (DC)

The DC process can be seen as a similar correction to Compton scattering as Bremsstrahlung is to Coulomb scattering, i.e. the equivalent reaction with one more photon in the outgoing channel:

$$e(p) + \gamma(k) \longleftrightarrow e(p') + \gamma(k') + \gamma(k_2). \tag{2.8}$$

This process dominates in photon production over BR at early times ($z \gtrsim 5 \times 10^5$), mostly due to the enormous dominance of photon number over baryon number in the early Universe.

The full cross sections for various limiting cases were originally studied by Mandl & Skyrme (1952), and a convenient summary can be found in chapter 11 of Jauch & Rohrlich (1976). For recent work, and a discussion of the relevance to SD, see

Chluba et al. (2007). As seen in the previous sections it is extremely helpful to have computational packages which allow for fast yet accurate calculations. In the context of DC they are based on

$$C[f_\gamma]\Big|_{\mathrm{DC}} \approx \frac{16\alpha\pi^3\theta_\gamma^2}{45x^3}e^{-2x}g_{\mathrm{DC}}\Big[1 - f_\gamma(e^{x_\mathrm{e}} - 1)\Big], \tag{2.9}$$

where θ_γ is defined analogously to θ_e, and g_{DC} is the DC equivalent of the Gaunt factors g_{ff}. Notice again that a blackbody spectrum is an explicit equilibrium state for $T_\gamma = T_\mathrm{e}$. Strong temperature dependence is contained in the g_{DC} term, as discussed in Chluba et al. (2007), where they find a suppression of $\sim 10 - 20\%$ at higher temperatures - proving these higher order corrections are non-negligible effects for physics in the early Universe. More recent work can be found in Ravenni & Chluba (2020), where the capabilities of `DCpack` are outlined. This code allows accurate estimates of DC contributions in wide volumes of parameter space, similar to that discussed for BR.

2.2 Spectral shapes

Following the historical study of the CMB we will begin by discussing the blackbody spectrum, and build up to more general cases. This approach is additionally a well justified considering that the CMB resembles a blackbody to very high precision.

2.2.1 Blackbody spectrum

The blackbody spectrum can be understood as a gas of photons (bosons) following Bose-Einstein statistics. The occupation number is given by

$$f_{\mathrm{bb}}(x) = \frac{1}{\exp(x) - 1}, \tag{2.10}$$

where x is a normalised frequency $x = h\nu/k_\mathrm{B}T$. The characteristic -1 allows many photons to occupy the same energy level.

It is occasionally useful to relate these abstract distributions to observable quantities. To find the distribution of energies of such a gas we must multiply the energy in each state by the number of photons that can exist in a given state, and finally by the density of states. With an explicit factor of 2 for the two polarisations of a photon, the density of states can be found as

$$\begin{aligned} \frac{df}{dk} &= 2V\frac{k^2}{2\pi^2}, \\ \frac{df}{d\epsilon} &= 2V\frac{\epsilon^2}{2\pi^2(\hbar c)^3}, \end{aligned} \tag{2.11}$$

where k and ϵ are the wavenumber and energy of an individual state. A factor of volume, V, ensures that all quantities remain intensive. The energy density of the gas, $u(x)$, is then given by

$$u(x) = \frac{8\pi(k_{\rm B}T)^3}{h^2c^3}\frac{x^3}{\exp(x)-1} = \frac{8\pi(k_{\rm B}T)^3}{h^2c^3}x^3 f_{\rm bb}(x) = \frac{4\pi}{c}I(x), \tag{2.12}$$

where we also defined the intensity $I(x)$. It is common to see this given per steradian on the sky, for which we would simply divide by 4π.

In this thesis we will not often link to directly observable quantities, but instead prefer dimensionless counterparts. In general we find the number density of a distribution $f(x)$ with

$$N = \int {\rm d}x\, x^2 f(x), \tag{2.13}$$

and the energy density as

$$E = \int {\rm d}x\, x^3 f(x). \tag{2.14}$$

It is useful to have these quantities in mind to compare with other spectral shapes:

$$N_{f_{\rm bb}} = 2\zeta(3) \approx 2.40411, \qquad E_{f_{\rm bb}} = \frac{\pi^4}{15} \approx 6.49394 \tag{2.15}$$

The CMB sky is approximated extremely well by just specifying its closest corresponding blackbody spectrum as seen in Fig. 2.1, thus providing a background temperature T. This will frequently be referred to as the CMB monopole temperature, a

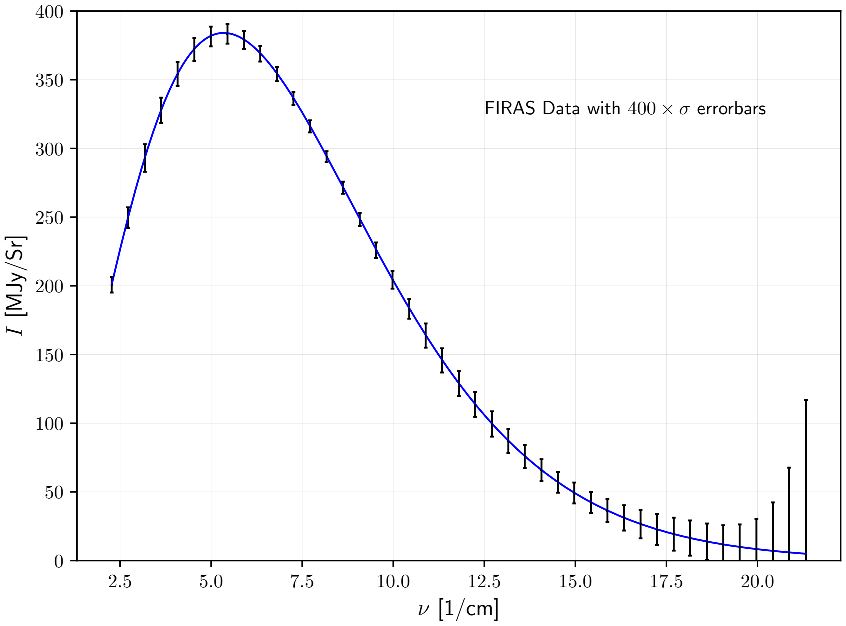


Figure 2.1: A figure showing the COBE/FIRAS data for the average photon spectrum seen across the sky. The error bars are multiplied by a factor of 400*, emphasising just how precise the spectral shape is. Data courtesy of NASA.*

term which will be clearer after discussing temperature shifts.

2.2.2 Temperature shifts

In the previous chapter we introduced the CMB anisotropies as small variations in temperature across the sky. Mathematically the temperature measured from the Earth now depends on observation angle of the sky $T \to T(\hat{\gamma})$. We also highlighted however that these fluctuations are very small, of order $\pm 300\mu$K compared to the background $T_0 = 2.725$K (Planck Collaboration, 2018a; Fixsen et al., 1996). This then motivates an expression where a perutrbatively small angle dependent quantity is superposed with the constant background temperature: $T(\hat{\gamma}) = \bar{T} + \delta T(\hat{\gamma}) = \bar{T}(1 + \delta T(\hat{\gamma})/\bar{T}) = \bar{T}(1 + \Theta(\hat{\gamma}))$. From now on we will drop the overbar notation, and always use T for the background temperature and Θ for the small angle dependent variations.

A key point here is that we can find a distribution which describes this small shift in temperature by Taylor expansion around some reference temperature and propagating that into the appropriate change of $x \propto T^{-1}$. We refer to this distribution as the temperature shift $\mathcal{G}(x)$:

$$\begin{aligned} f(x) &\approx f(x') + \left.\frac{\partial f_{\rm bb}}{\partial T}\right|_{T_{\rm av}} \delta T \\ &= f(x') - x' \left.\frac{\partial f_{\rm bb}}{\partial x}\right|_{T_{\rm av}} \Theta \\ &= f(x') + \mathcal{G}(x')\Theta. \end{aligned} \tag{2.16}$$

Some useful relations for temperature shifts to remember are

$$\mathcal{G}(x) = -x\partial_x f_{\rm bb} = x f_{\rm bb}(f_{\rm bb} + 1) = \frac{1}{4} x \,{\rm csch}^2(x/2). \tag{2.17}$$

The way this shape alters the background blackbody spectrum is shown in Fig. 2.2. Also shown in dotted lines is the result of simply evaluating a blackbody distribution, showing that small differences appear for ~ 10% variations. Temperature shifts of this size are orders of magnitude larger than the real variations seen across the sky. In Fig. 2.3 in contrast we have subtracted the reference temperature which firstly highlights the shape of a pure T-shift, and secondly allows us to show realistic size of fluctuations at ~ 0.01%, showing the first order shift to be valid. This process of studying only a relative change is more akin to how calculations are performed. As such, figures with peaks and valleys showing relative excess and absence of photons will become familiar as we explore the spectrum more deeply.

One important aspect of temperature shift spectra is that they carry photon number as well as energy:

$$N_{\mathcal{G}} = 6\zeta(3) = 3N_{f_{\rm bb}} \approx 7.21234, \qquad E_{\mathcal{G}} = \frac{4\pi^4}{15} = 4E_{f_{\rm bb}} \approx 25.9758. \tag{2.18}$$

This becomes important in defining other spectral shapes.

Note that here we neglected terms $\mathcal{O}(\Theta^2)$ and higher since our interest is just the first order temperature shift. Below we will perform a similar calculation to higher order to study the mixing of blackbody spectra.

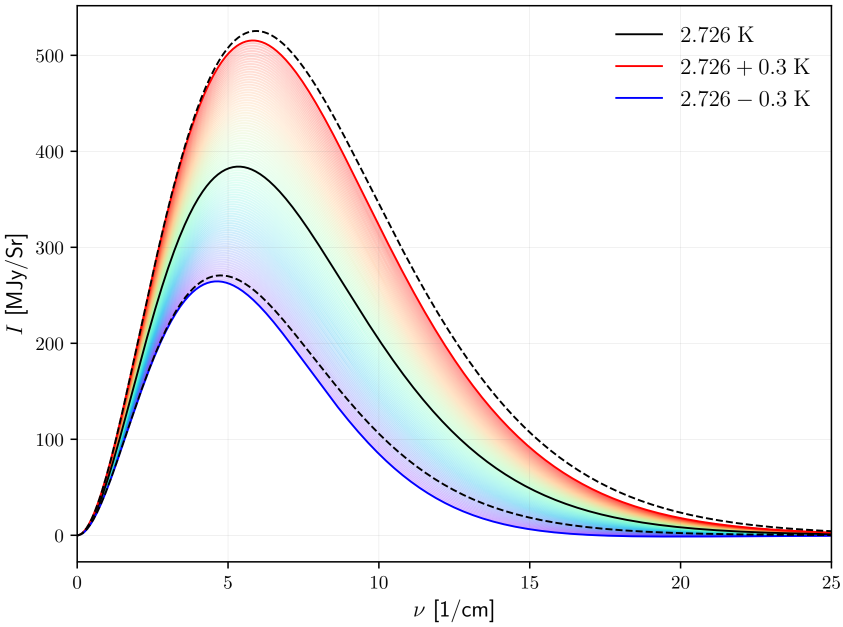


Figure 2.2: A plot showing how temperature shifts added to a blackbody simply shift the blackbody in predictable ways. e.g. the peak of the distribution shifts to higher frequencies and amplitudes with higher temperature. Dashed lines show the result of evaluating the blackbody distribution at the new temperature, revealing small innacuracies for 10% *variations of temperature, which would be remedied by adding y-distortions (defined below).*

Here we are simply leaving the variables as a function of direction vector $\hat{\gamma}$. In studying the CMB sky however we analyse them in a spherical basis $\Theta_{\ell m}$. Note however that Θ_{00} would only shift the temperature T_0, and hence $\Theta_{00} = 0$ by definition – hence the statement above that the average temperature is known as the monopole temperature. Similarly Θ_{1m} is dependent on the relative velocity of the measurement, and is subtracted in analysing the CMB.

2.2.3 y-Distortions

The previous section derived a spectral shape that to first order leaves a blackbody unchanged. This resulted in a statement that the monopole of temperature shifts are either 0 or simply ill defined, depending on your outlook. In this section we will take the first step towards a completely new spectral shape, that would be truly distinguishable

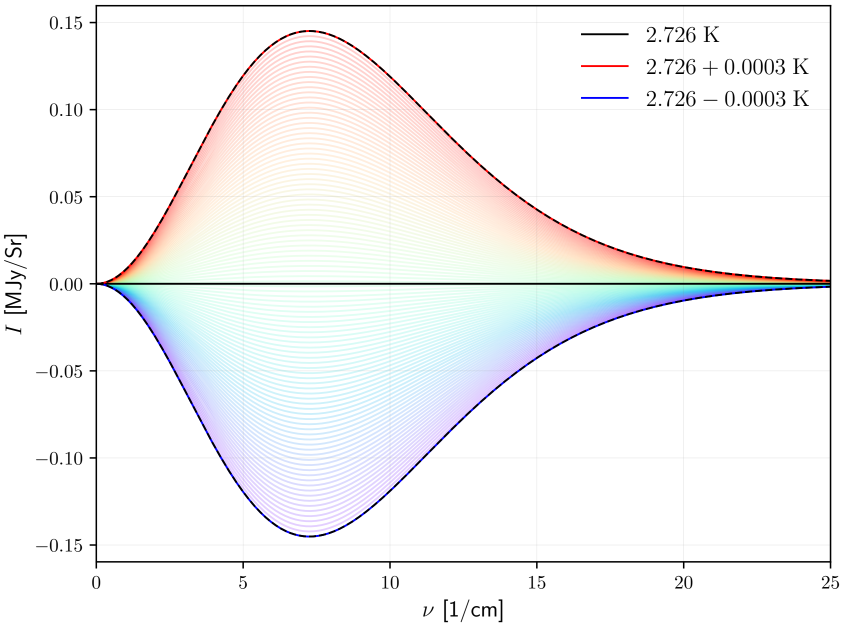


Figure 2.3: A plot showing the relative changes in photon spectrum captured in the temperature shift spectrum $\mathcal{G}$. Here a negative value indicates a relative lack of photons, and a positive branch shows an excess. The amplitude of variations here are comparable to that seen in real CMB data. The dashed lines again show an exact evaluation, showing here that temperature shifts are sufficient to model these small variations.

from a blackbody. We refer to this as a y distortion, since it shares its shape with another astrophysical signal, the Sunyaev Zeldovich effect (Zeldovich & Sunyaev, 1969).

Stated in a heuristic way, the y distortion emerges in many different aspects of cosmology and astrophysics simply because it is quite a fundamental spectral shape. Consider for example the spectral shape produced upon scattering a blackbody with an electron gas of a different temperature. This scenario is captured via the Kompaneets equation (Kompaneets, 1957), and gives

$$\begin{aligned}
\Delta f &\approx \Delta\tau \frac{\theta_{\rm e}}{x^2}\frac{\partial}{\partial x} x^4 \Big(\frac{\partial}{\partial x} f_\gamma + \frac{T_\gamma}{T_{\rm e}} f_\gamma [1 + f_\gamma]\Big) \\
&\approx \Delta\tau \frac{\theta_{\rm e}}{x^2}\frac{\partial}{\partial x} x^4 \Big(\frac{\partial}{\partial x} f_{\rm bb} + \frac{T_\gamma}{T_{\rm e}} f_{\rm bb} [1 + f_{\rm bb}]\Big) \\
&= \Delta\tau \frac{\theta_{\rm e}}{x^2}\frac{\partial}{\partial x} x^4 \Big(\frac{T_\gamma}{T_{\rm e}} - 1\Big) f_{\rm bb}(f_{\rm bb} + 1) \\
&= \Delta\tau x^{-2}(\theta_\gamma - \theta_{\rm e}) \frac{\partial}{\partial x} x^4 f_{\rm bb}(f_{\rm bb}+1) \\
&= \Delta\tau (\theta_\gamma - \theta_{\rm e}) x f_{\rm bb}(f_{\rm bb}+1) \Big[4 - (2 f_{\rm bb} + 1)x\Big] \\
&= \Delta\tau (\theta_\gamma - \theta_{\rm e}) \mathcal{G}(x) \Big[x \coth(x/2) - 4\Big].
\end{aligned} \tag{2.19}$$

From here we define

$$\mathcal{Y}_{\rm SZ}(x) = \mathcal{G}(x)\Big[x\coth(x/2) - 4\Big], \tag{2.20}$$

with an amplitude of

$$y = \Delta\tau(\theta_\gamma - \theta_{\rm e}). \tag{2.21}$$

A second scenario where these distortions arise together with a temperature shift is in the mixing of two blackbody spectra at different temperatures - a simple process that turns out to capture a lot of the physics behind the SD caused by dissipation of acoustic waves (Chluba et al., 2012). First let us expand one blackbody spectrum at T around a different temperature $T_{\rm av}$ as we did earlier, but now to second order:

$$\begin{aligned}
f(x) &= \frac{1}{e^x - 1} = \frac{1}{e^{x_{av}/[1+\Theta]} - 1} \\
&= f_{\rm bb}\Big|_{T_{\rm av}} + \frac{\partial f_{\rm bb}}{\partial T}\Big|_{T_{\rm av}} \delta T + \frac{1}{2}\frac{\partial^2 f_{\rm bb}}{\partial T^2}\Big|_{T_{\rm av}} \delta T^2 + O(\delta T^3),
\end{aligned} \tag{2.22}$$

where $T = T_{\rm av} + \delta T = T_{\rm av}[1 + \Theta]$ and $x_{\rm av} = \frac{h\nu}{k_{\rm B} T_{\rm av}}$. By then using the chain rule $\frac{\partial f}{\partial T} = \frac{\partial x}{\partial T}\frac{\partial f}{\partial x}$ it can be found that

$$f(x_{\rm av}) \approx f_{\rm bb}(x_{\rm av}) + \mathcal{G}(x_{\rm av})(\Theta + \Theta^2) + \frac{1}{2}\mathcal{Y}_{SZ}(x_{\rm av})\Theta^2. \tag{2.23}$$

When combining two blackbody spectra at temperatures $T_{\rm av} \pm \frac{\delta T}{2}$ the first order contribution will cancel, leaving an overall change $\mathcal{G} + \frac{1}{2}\mathcal{Y}_{SZ}$ at second order.

In the previous derivation we separated the second order term into a temperature shift and a y distortion, a step which would not have been obvious unless we knew what we were looking for in advance. The y distortion shape could have been derived more agnostically by simply performing the second order expansion, and then subtracting the necessary amount of $\mathcal{G}(x)$ till the distortion carried no photon number. This demand that distortions do not add or subtract photon number is a useful choice, since it allows us to focus on energy contributions with less ambiguity (e.g. see chapter 6). The number and energy properties of the y distortion are found to be

$$N_{\mathcal{Y}_{\rm SZ}} = 0, \qquad\qquad E_{\mathcal{Y}_{\rm SZ}} = \frac{4\pi^4}{15} = 4E_{f_{\rm bb}} \approx 25.9758. \tag{2.24}$$

A visual description of this process is shown in Fig. 2.4, where it can be seen that mixing two blackbody leaves some distortion in the high frequency part of the distribution, known as the *Wien tail*. The low frequency part of the distribution, known as the *Rayleigh-Jeans* (RJ) *tail*, remains unchanged. In the following section we will see another distortion shape which affects the low frequencies instead. The relative shape of the y distortion (once the reference blackbody is subtracted) can be seen in Fig. 2.5 (now in dimensionless units, as for the rest of this thesis).

An interesting consequence of the emergence of $\mathcal{Y}_{\rm SZ}$ from blackbody mixing is that this can occur either by real mixing of fluids at different temperatures, or by measuring a patch of the sky containing various temperatures. The utility of the latter case, the *beam spectral distortion*, was discussed in Chluba & Sunyaev (2004).

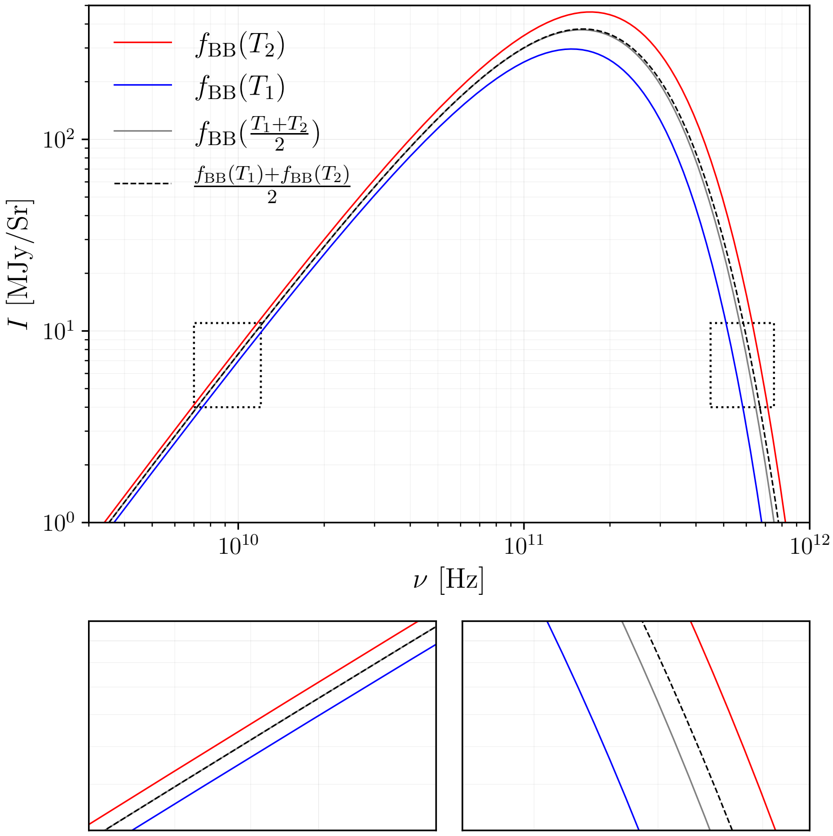


Figure 2.4: A figure showing how the superposition of two blackbody spectra does not give another blackbody, rather there will inevitably be some SD. In the top panel the blue and red lines show the blackbody spectra at two different temperatures, and the dotted line shows the blackbody at the average temperature between them. In the bottom panels the resulting mixed spectrum is shown, demonstrating that it follows the dotted line till the Wien tail, where a distortion is present. Figure adapted from Chluba et al. (2015b).

2.2.4 μ-Distortions

When introducing Eq. (2.10), we did not include a chemical potential $\tilde{\mu}$, despite a general Bose-Einstein distribution being

$$f_{\rm BE}(\nu) = \frac{1}{\exp\left(\frac{h\nu-\tilde{\mu}}{k_{\rm B}T}\right)-1}. \tag{2.25}$$

Why did we initially neglect the potential extra term? $\tilde{\mu}$ is associated with an *energy cost* of having a particle in your system, e.g. rest energy of massive particles. In the

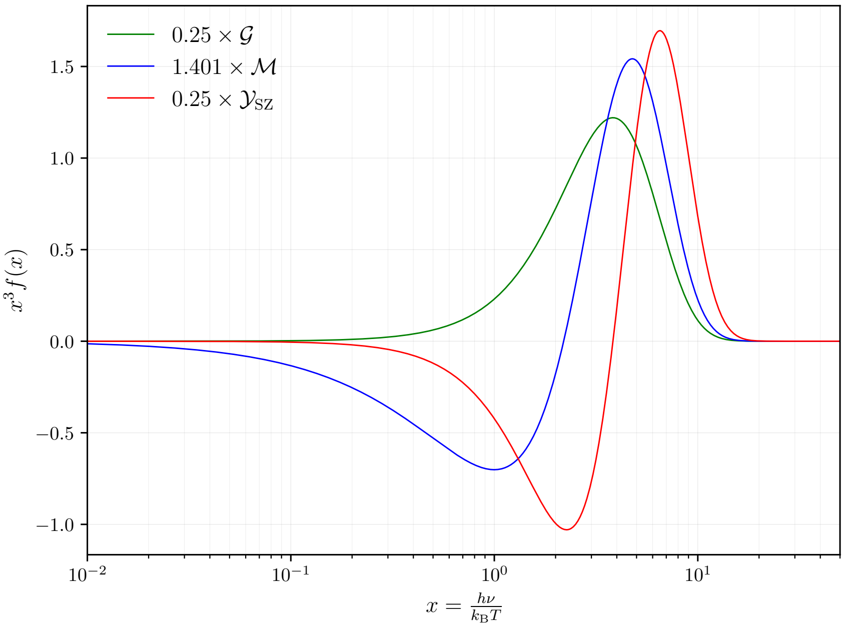


Figure 2.5: A plot visualising the three main spectral distortions as relative spectra. This means the negative branches show relative absence of photons while the positive branches are excess photons. Roughly speaking y distortions push photons from the blackbody peak into the high frequency tail, while μ distortions push photons from the low frequency tail towards the blackbody peak.

case of photons they are massless, and can in principle be created/destroyed *for free*. There are times, however, in the Universe's history where no mechanisms for photon creation/annihilation were active, which creates an effective chemical potential. Let us see this mathematically by analysing the equilibrium solution of the Kompaneets equation.

$$\begin{aligned} &\frac{df_\gamma}{dt} = 0, \\ \implies &\frac{\partial}{\partial x} f_\gamma + \frac{T_\gamma}{T_{\rm e}} f_\gamma [1 + f_\gamma] = 0, \end{aligned} \tag{2.26}$$

A useful trick here is to recast the dimensionless frequency x in terms of electron temperature

$$x = \frac{h\nu}{k_{\rm B}T_\gamma} = \frac{T_{\rm e}}{T_\gamma}\frac{h\nu}{k_{\rm B}T_{\rm e}} = \frac{T_{\rm e}}{T_\gamma} x_{\rm e}. \tag{2.27}$$

Using this definition we can express Eq. 2.26 as

$$\partial_{x_\mathrm{e}} f_\gamma = -f_\gamma(1 + f_\gamma), \tag{2.28}$$

which is solved by a blackbody with an extra term μ given by

$$f_\gamma = \frac{1}{\exp\left(x_\mathrm{e} + \mu\right) - 1}. \tag{2.29}$$

This is analogous to the chemical potential, but is not defined exactly the same (notice the difference in sign and units, and hence the intentional change of symbol). To describe the spectral shape of the μ distortion, $\mathcal{M}(x)$, we assume the spectrum is close to a blackbody, and an expansion in small μ reveals

$$f_\gamma(x_\mathrm{e}) \approx f_\mathrm{bb}(x_\mathrm{e}) - \mu \frac{\mathcal{G}(x_\mathrm{e})}{x_\mathrm{e}}. \tag{2.30}$$

As mentioned earlier we will study distortions which leave the photon number unchanged, so we again subtract $\mathcal{G}(x)$ with the appropriate prefactor to achieve the final form of $\mathcal{M}(x)$:

$$\mathcal{M}(x) = \mathcal{G}(x)\left(\alpha_\mu - \frac{1}{x}\right), \tag{2.31}$$

with $\alpha_\mu \approx 0.4561$. This leaves the properties

$$N_\mathcal{M} = 0, \qquad E_\mathcal{M} = \frac{2\pi^6}{135\zeta(3)} - 6\zeta(3) \approx \frac{E_{f_\mathrm{bb}}}{1.40066} \approx 4.63635. \tag{2.32}$$

For a more detailed study of the μ distortion shape, origin and evolution see Chluba (2014a). This shape is shown in Fig. 2.5, where it is seen to subtract photons from the low frequency RJ tail and push them towards the peak of the underlying blackbody distribution.

2.3 Primordial origins of SDs

So far we have introduced a variety of spectral shapes with motivating arguments for each. In this section we will piece these arguments together to form a coherent view of which SD probes which era in cosmological history. In stating this we are already anticipating part of the punchline: observing SD allow us to trace events in the primordial plasma behind the CMB curtain. Different spectral shapes reveal a different era, and thus we will build up the three era picture. A useful concept to have in mind throughout the rest of this chapter is that the true blackbody is akin to the thermal equilibrium for a photon gas. The gas will strive to reach this equilibrium state, however it must undergo interactions to do so. The availability of these interactions govern the three eras (Lucca et al., 2019; Chluba, 2016).

To quantify these eras it is helpful to introduce the energy branching ratio $\mathcal{J}_{\rm x}(z)$, which specifies the amount of energy that will be locked into a spectral shape of type x following energy injection at redshift z. Denoting the relative energy injection to the photon gas as Q/ρ_γ, we can perform simple calculations of SD amplitudes by evaluating

$$y = \frac{1}{4}\int {\rm d}z\, \mathcal{J}_y(z)\frac{{\rm d}(Q/\rho)}{{\rm d}z}, \tag{2.33}$$

$$\mu = 1.401\int {\rm d}z\, \mathcal{J}_\mu(z)\frac{{\rm d}(Q/\rho)}{{\rm d}z}, \tag{2.34}$$

where the prefactors balance the amplitudes in terms of energy content of their corresponding spectral shape. These branching ratios are shown in Fig. 2.6, together with shading indicating the rough separation of the three eras. These lines are drawn using “Method C” from Chluba (2016).

2.3.1 The Temperature Era

This era is the earliest in history, and is characterised by a universe much smaller than today, and thus very energetic. Energy exchanging scattering events between electrons and photons occur with great frequency. Bremsstrahlung and Double Compton emission

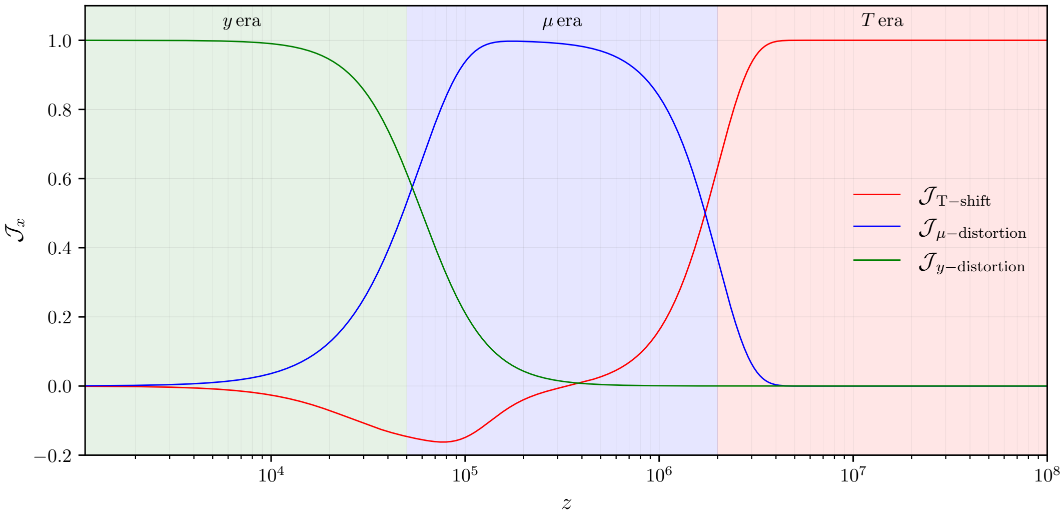


Figure 2.6: A plot showing the three era picture with energy branching ratios (Chluba, 2016). Deep into each era the corresponding branching ratio tends to unity, indicating the production of a single spectral shape from energy injection at that time.

are still efficient at these times, and any perturbations in the photon gas' spectral shape will inevitably return to a pure temperature shift, thus the naming of the era. This era lasts till $z \approx 3 \times 10^6$. This boundary however is sensitive to exactly how distorted the spectrum is initially, with large energy injections being potentially observable at earlier times (Chluba et al., 2020a).

2.3.2 The μ Era

As the Universe expands and lowers its average energy density the events associated with photon creation become more and more unlikely. The scattering events are still happening however, meaning there is still an up scattering of photons upon interaction with hot electrons. This leads to a situation where low frequency photons have been scattered up towards the peak of the distribution, but not enough soft photons are created to compensate that. The result is the μ distortion, with its absence of low frequency photons and its effective non-zero chemical potential. This era lasts until $z \approx 5 \times 10^4$.

We note that given the power law evolution we saw for matter and radiation in Sect. 1.3 we would naturally expect electrons to be colder than photons. In reality however any energy injection into the medium will tend to heat electrons before transferring energy to photons. Even in the standard ΛCDM cosmology this effect is larger than the Hubble cooling effect on matter, at least in the primordial plasma (we discuss this ΛCDM energy source in the following chapter).

2.3.3 The y Era

The era that follows is the logical extension of the fact that an expanding and cooling universe will eventually host almost no interactions. In this era the energy exchanging scattering events that allow a y distortion to convert into a μ distortion are no longer efficient. A question then arises as to the name of this era: it may not be possible to thermalise a y spectral shape, but why do we hypothesise a y shape in the first place. Generally speaking a y distortion emerges as characteristic shape following most forms of energy injection. To see this we will allude to our two mechanisms which served to introduce and derive the shape in Sect. 2.2.3. Firstly there will be a y distortion any time the electrons are out of equilibrium with photons, which is expected even just from expansion history. Importantly, however, electrons interact with other components of the primordial plasma before communicating to photons through electromagnetic forces. This means any general mechanism for energy injection which heats matter will be translated to a y spectral shape. Secondly, we again note the superposition of blackbodies which constitutes an energy injection in a real sense, and also leave a y shape. This second mechanism is especially important in the early Universe, since the growing mean free path of photons as the medium becomes less dense naturally yields this mixing (Silk, 1968).

2.3.4 The residual era

It may be clear that the eras discussed so far are all simple limiting cases involving assumptions about available thermalization mechanisms and their respective efficiencies. In reality the picture is more complicated, especially between the μ and y eras where the transition happens with some finite time. This leads to the existence of a residual era, where intermediate spectral shapes must be described. These shapes would carry more sensitivity to the precise moment of energy release in the Universe's history. One method of decomposing the residual distortion into orthogonal components has been developed (Chluba & Jeong, 2014), and a Green's function method which makes use of that formalism (Chluba, 2013, 2015). While no simple formula exists describing the broad class of shapes composing residual distortions, some approach is made in this direction in chapter 6.

2.4 Spatial spectral distortions

This endeavour of understanding the photon spectrum in depth was motivated by the fact that photons constitute almost the entirety of cosmological information at our disposal. Here we will double down on this sentiment by saying that observing the CMB sky has almost independently informed our models when it comes to early Universe cosmology in particular. By this we mean there are no other early light sources considering that stars and galaxies were absent for many billions of years still, and the other probes we mentioned usually arise in the late Universe [most gravitational wave events for example arise from neutron star black hole merges (Abbott et al., 2019)].

An argument can be made then that it is imperative to extract and dissect all the information possible from the CMB sky. This on the one hand means searching for spatial information like the CMB temperature anisotropies [$\Theta(\hat{\gamma})$], but on the other hand it means searching for information locked in the frequency distribution beyond a blackbody shape [$y\mathcal{Y}_{\rm SZ}(x) + \mu\mathcal{M}(x)$]. The former give us rich information about the universe very close to recombination, while the latter push our frontier of knowledge further back in time.

One thing we didn't discuss here is the possibility of spatially varying SD [$y(\hat{\gamma})$, $\mu(\hat{\gamma})$], although it will be the main topic of discussion in chapters 6-8. We can assert at this stage however that a future decomposition of the CMB sky in exquisite detail could follow

$$f_\gamma(x, \hat{\gamma}) = T_{\rm CMB} f_{\rm bb}(x) + \Theta(\hat{\gamma})\mathcal{G}(x) + y(\hat{\gamma})\mathcal{Y}_{\rm SZ}(x) + \mu(\hat{\gamma})\mathcal{M}(x), \tag{2.35}$$

where all symbols take the meaning we have introduced in this chapter, but now with the inclusion of spatially varying y and μ amplitudes. It is expected that these signals would be some orders of magnitude smaller than the monopole SD, but would nevertheless provide an even richer information content than already expressed throughout this chapter. A further benefit of searching for anisotropic signals is that we can benefit from the wealth of mathematical and computational machinery already developed for the extraction of temperature anisotropies – highly optimised codes (Lesgourgues, 2011; Lewis et al., 2000) and models of foregrounds (Rotti & Chluba, 2021; Remazeilles & Chluba, 2018), just to mention two.

3

Cosmological Perturbation Theory

3.1 Introduction

In the previous chapters we have introduced the standard model of cosmology and discussed the information content locked in the photon spectrum we see across the sky. This chapter will both develop and unite the two previous concepts. To see this consider the photon decomposition represented in (2.35), where a large portion of the information content manifests in angular dependence across the sky. This is at odds with the discussion in the first chapter, where the Universe is hypothesised to be maximally symmetric - homogeneous and isotropic - and as such any angular dependence of photons would be impossible. Similarly, it has been extremely helpful while writing this thesis to reside upon a dense ball of rock floating in space, a fact which is also prohibited by a maximally symmetric universe, as no location is preferable for gravitational collapse. Observations have thankfully shown our Cosmological models to not be completely unjustified, as differences in temperature across the sky $\Theta(\hat{\gamma})$ are typically only 1 part in 100 thousand of the monopole temperature. This motivates us to take a perturbative approach, keeping a maximally symmetric Universe at leading order.

To formally treat this problem we will introduce the study of Cosmological Perturbation Theory (CPT) where perturbatively small inhomogeneities and anisotropies are added to the otherwise smooth background. This formalism allows us to fully explain the sizes of temperature patches on the CMB sky by calculating the power spectrum of CMB anisotropies. Additionally we will make connection to SD production by explaining the dominant source of primordial SDs within ΛCDM - Silk damping of acoustic modes.

We will see throughout this chapter though that CPT can be fiercely complicated. Let us see why in a heuristic way: the early-universe cosmic inventory contains charged baryons and electrons, dark matter, photons, neutrinos and gravitational potentials. All of these oscillate as a function of both space and time. Since the matter can be considered non-relativistic they are each expressed with densities δ_X and velocities v_X, where X can denote charged baryonic[1] matter (e.g. $\delta_{\rm b}$) or uncharged dark matter (e.g. $\delta_{\rm m}$). The neutrinos and photons however must be treated as more general fluids, but cannot be assumed perfect fluids (see Sect. 1.3), and as such have hierarchies of multipoles $X_{\ell m}$. The gravitational potentials contain 6 degrees of freedom divided into scalars, vectors and tensors (see Sect. 3.3). The charged particles and photons interact through familiar scattering terms (see Sect. 2.1), but even the weakly interacting particles like neutrinos will couple to the entire system through the gravitational potentials. We will see that this enormously coupled and energetic system of interactions can be distilled down to coupled ODEs, albeit potentially hundreds of them.

This section is split in two parts, both of which rely on each other to some degree. Firstly, we must take a closer look at the Einstein field equations Eq. (1.11), which stated in words dictates “how space time should curve in response to stress-energy content”. In this perturbative approach then this equations tells us how small local potentials are created on top of the global FLRW geometry according to the perturbed energy content. Secondly, we must look at is the other half of the Einsteinian world view where “spacetime curvature tells particles how to move”, where we rely on the gravitational potentials to solve this. One complication however arises in the aforementioned particle interactions, which can be seen generally as terms that modify the phase space distribution of particle species (e.g. the occupation number we discussed for photons in Sect. 2.2). This second problem can be formulated through the Boltzmann equation – a deceptively simple statement of phase space evolution with appropriate collision terms – where we can insert the known evolution across a spacetime geodesic given the geometry and gravitational potentials.

---

[1] Recall from chapter 1 that this means both electrons and protons in the context of cosmology.

3.2 The Boltzmann Equation

The Boltzmann equation can be given in a deceptively simple form (Peebles & Yu, 1970; Weinberg, 2008; Dodelson, 2003):

$$\frac{\mathrm{d}f_i(\boldsymbol{x},\boldsymbol{p},t)}{\mathrm{d}t} = C[f(\boldsymbol{x},\boldsymbol{p},t)], \tag{3.1}$$

where f_i is the occupation function, essentially a function describing the distribution of a particle species i in phase space. The phase space is parametrised in terms of position $\boldsymbol{x}$ and momentum $\boldsymbol{p}$. Most important macroscopic physical quantities of the particle baths are found by taking various integrals over the occupation function. We already

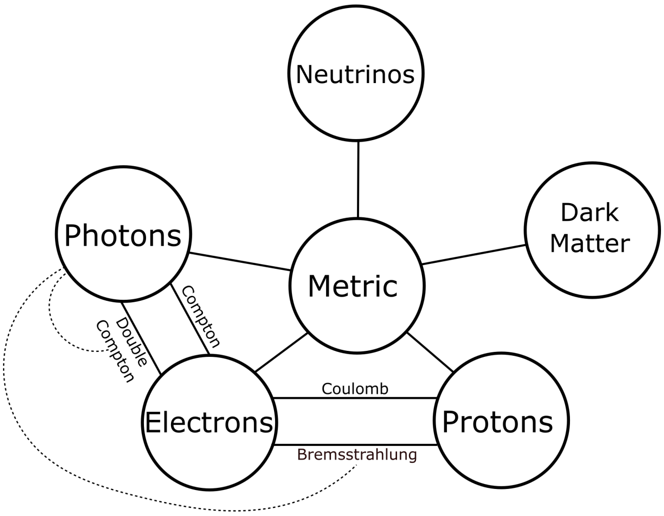


Figure 3.1: A diagram adapted from Dodelson (2003) showing different components in the early Universe with relevant interactions. Dotted lines indicate photons being produced or absorbed in the reaction. All lines connecting to the metric are understood as gravitational interactions.

discussed the photon occupation number in depth (see Sect. 2.2).

The left and right hand side of Eq. (3.1) are known as the Liouville and collision operator respectively. In essence the Boltzmann equation is a statement of how the occupation number will change under various interactions: injection/absorption at some frequency, scattering up/down in frequency, diffusions in space, etc. As such, they tend to be functions of the momenta and they often couple different particle types. This is diagrammatically summarised in Fig. 3.1, where circles indicate elements of the cosmic inventory (and the metric, to clarify the coupling of gravity), with vertices indicating various interactions. We will not discuss collision terms in any depth in this chapter, since we already covered the important ones for photons in Sect. 2.1.

3.2.1 The Liouville Operator

Given the structure of the Boltzmann equation we can identify the Liouville operator as containing the evolution of the occupation function in the absence of any sources/scatterings. This means it is what particles do following their "straight lines" through spacetime. In the context of General Relativity this is less boring than it sounds, involving the loss of momentum to cosmic expansion and curved paths from local gravitational masses. To see this let us expand the operator and reveal its physical content:

$$\frac{\mathrm{d}f_\gamma}{\mathrm{d}t} = \frac{\partial f_\gamma}{\partial t} + \frac{\partial f_\gamma}{\partial x^i}\frac{\mathrm{d}x^i}{\mathrm{d}t} + \frac{\partial f_\gamma}{\partial p^i}\frac{\mathrm{d}p^i}{\mathrm{d}t}, \tag{3.2}$$

Studying the terms from left to right we can see explicit dependence on time, a dependence on the way physical coordinates vary through time, and any change in the momentum of particle (e.g. redshifting photons). It is clear then that the geometry of spacetime will enter the calculation through non-trivial changes to the coordinates and momentum.

Eq. (3.2) has been written using momentum p, however for the rest of this thesis it will be more helpful to express equations using the dimensionless frequency $x = h\nu/k_\mathrm{B}T$ as defined in Chapter 2. This variable which has the Cosmic expansion *built in* through

$T = T_0/a = T_0(1+z)$, and thus absorbs a term which would otherwise appear giving that precise temperature scaling (Dodelson, 2003). Throughout the rest of this section we use the notation from Chluba et al. (2012) (Chapter 7 follows the same convention).

Upon the introduction of perturbations, the geometry of spacetime has to include any gravitational effects of under- and over densities in the Universe. This will be discussed in detail in section 3.3, but for now all we need to know is that two gravitational potentials, Φ, Ψ, cause this effect (see Eq. (3.25)). The Liouville operator then takes the form (Hu & White, 1996; Chluba et al., 2012)

$$\frac{\mathrm{d}f^{(1)}}{\mathrm{d}t} = \frac{\partial f^{(1)}}{\partial t} + \frac{\hat{\gamma}^j}{a}\frac{\partial f^{(1)}}{\partial x^j} - x\frac{\partial f^{(0)}}{\partial x}\left[\frac{\partial \Phi^{(1)}}{\partial t} + \frac{\hat{\gamma}^j}{a}\frac{\partial \Psi^{(1)}}{\partial x^j}\right], \tag{3.3}$$

where the effects of magnitude and direction of momentum have been explicitly separated by introducing $\hat{\gamma}^i = p^i/\sqrt{p^j p_j}$ and we have dropped the subscript γ, understanding we are interested only in photons. Note that here we are now specifically studying first order perturbations in preparations for studying anisotropies (additionally we used that background quantities don't depend on spatial coordinates). Inspecting Eq. (3.3) reveals the interesting property that *all* fluids will interact with the geometric perturbations via the Liouville operator. This explains the vertices in Fig. 3.1, which connect every fluid to the metric (e.g. even neutrinos will have small effects on the photons via gravity, despite having no direct electromagnetic interaction).

The first two terms in Eq. 3.3 are the generalised forms of the continuity and Euler equations. The next term provides the loss of energy to cosmic expansion – cosmic redshifting (the equivalent at zeroth order reveals $T \propto a^{-1}$). The final terms are related to gravitational lensing – they describe the change of frequency and angle in being deflected by local potentials.

3.2.2 The collision operator

The collision operators will not be discussed in great detail here simply due to the fact that they were discussed at length in Sect 2.1, albeit in a different context. Due to that context we previously focused on energy exchanging events between photons and electrons, as well as photon creating events. In this chapter we in fact care much more about the low energy equivalent of Compton scattering, the Thomson limit. This can be given by

$$C^{(1)}[f(x)] = \dot{\tau}\left[f_0^{(1)} + \frac{1}{10}f_2^{(1)} - f^{(1)} - \beta^{(1)}\chi x\partial_x f^{(0)}\right], \tag{3.4}$$

where $\beta^{(1)}$ represents local baryon velocities, and χ is the angle between the photons ($\hat{\gamma}^j$) and baryons. The operator $x\partial_x$ appears frequently, and is explicitly named the boost operator in Chapters 6, 7 and 8. The subscripts of f are referring to the multipole moment of the distribution, where we have defined $f_\ell = \sum_{m=-\ell}^{\ell} f_{\ell m}Y_{\ell m}$ (see appendix A).

3.2.3 The Photon hierarchy

The bottom line of the calculations alluded to above is what we call the Boltzmann hierarchy, and in particular we will focus on the photon hierarchy provided that in this thesis we are primarily concerned with the photon fluid. Much of the above discussion holds for the rest of the Cosmic inventory however (excluding the collision term Eq. 3.4 which was specific for photons). Thankfully the other fluids are simple in the early universe: following the discussion in Sect. 1.3 we treat matter (CDM and baryons) as perfect fluids with only local densities and velocities as parameters. Photons and neutrinos on the other hand receive a full non-perfect fluid description This restriction of focus is not as limiting as it may initially seem, considering that the most important interactions in the plasma are electromagnetic, since gravity is included implicitly through the geometric view of GR.

The bottom line equation for photons in the Thomson limit is given by combing Eqs. (3.3) and (3.4),

$$f^{(1)\prime} + \hat{\gamma}^i\frac{\partial f^{(1)}}{\partial x^i} = -\Phi^{(1)\prime} - \hat{\gamma}^i\frac{\partial \Psi^{(1)}}{\partial x^i} - \tau'\left[f_0^{(1)} + \frac{1}{10}f_2^{(1)} - f^{(1)} - \beta^{(1)}\chi x\partial_x f^{(0)}\right], \tag{3.5}$$

where we have also converted to conformal time ($\prime \equiv \partial_\eta$). This will more often be seen in the literature with a "temperature only" view of the photon spectrum: $f^{(0)} \rightarrow f_{\rm bb}(x)$, $f^{(1)} \rightarrow \Theta^{(1)}\mathcal{G}(x)$. Importantly using that substitution has an internal consistency in that $x\partial_x f_{\rm bb}(x) = \mathcal{G}(x)$. This mapping to only temperature shifts is at odds with the nuances of chapter 2, a discrepancy we attempt to remedy in chapter 7.

To understand why Eq. 3.5 makes part of a "hierarchy" we must first perform two mathematical transformations (appendix A). They both come down to a similar concept of taking our currently complicated function which depends on space and projects that dependence onto some set of basis functions. In particular the angular dependence of variables is captured through a hierarchy of Legendre polynomials $\mathcal{P}_\ell(x)$ (A.1) and length scales captured through the plane waves $\exp(ikx)$ (A.3) implicit in Fourier transforms. Transformations on the hierarchy are covered with great clarity in Pettinari (2016).

Bottom line

After performing all the relevant transformations we arrive at what can finally be recognised as a hierarchy:

$$\begin{aligned}
\hat{\Theta}'_0 &= -k\hat{\Theta}_1 - \hat{\Phi}', && (3.6)\\
\hat{\Theta}'_1 &= k\left(\frac{1}{3}\hat{\Theta}_0 - \frac{2}{3}\hat{\Theta}_2\right) + \frac{k}{3}\hat{\Psi} - \tau'\left[\hat{\Theta}_1 - \frac{\hat{\beta}}{3}\right], && (3.7)\\
\hat{\Theta}'_2 &= k\left(\frac{2}{5}\hat{\Theta}_1 - \frac{3}{5}\hat{\Theta}_3\right) - \frac{9}{10}\tau'\hat{\Theta}_2, && (3.8)\\
\hat{\Theta}'_{\ell>2} &= k\left(\frac{\ell}{2\ell+1}\hat{\Theta}_{\ell-1} - \frac{\ell+1}{2\ell+1}\hat{\Theta}_{\ell+1}\right) - \tau'\hat{\Theta}_\ell, && (3.9)
\end{aligned}$$

where *hats* denote the Legendre transforms of the variables introduced above (appendix A) and we drop the superscript (1) for convenience (Θ is always a perturbed quantity).

3.3 Einstein's field equations

Einstein's field equations are typically very difficult to solve analytically and costly to treat numerically. Full solutions tend to possess high degrees of symmetry, as seen in the FLRW case. In fact, FLRW corresponds to the most symmetric possible solution for the spatial part of the manifold. It can be shown (See appendix C of Wald (1984)) that a Riemannian manifold of dimension n has a maximum of $n(n+1)/2$ linearly independent Killing vector fields, which dictate the symmetries of the manifold. In 3 dimensional space the maximal number of Killing fields is $3(3+1)/2 = 6$, which in FLRW consists of 3 spatial translations (homogeneity) and 3 rotations (isotropy). This only leaves freedom of choice in the curvature (flat, open or closed), which we fix experimentally.

The use of perturbation theory is essential to move beyond the most symmetric cases and include, for example, the inhomogeneities that seed galaxies, stars and planets. Small linear perturbations on top of a simple background allow us to capture some of the less idealised gravitational effects, without an overwhelming analytic or numerical treatment. This is, fundamentally, the treatment that allows us to feasibly calculate the expected distribution of temperature patches across the sky. For a thorough review of these calculations see Weinberg (2008) (specifically chapter 5).

To perturb the geometry we consider a metric

$$g_{\mu\nu} = \bar{g}_{\mu\nu} + \delta g_{\mu\nu}, \tag{3.10}$$

where any background quantities are now represented by overbars. Any small quantities, indicated with δX, have small component values compared to the background. The raising and lowering of indices is no longer as simple as before. If we demand that $g_{\mu\alpha}g^{\alpha\nu} = \delta^{\nu}_{\mu}$ then it can be seen that

$$g^{\mu\nu} = \bar{g}^{\mu\nu} - \delta g^{\mu\nu} \tag{3.11}$$

is the correct form for the new inverse metric:

$$\begin{aligned} g_{\mu\alpha}g^{\alpha\nu} &= (\bar{g}_{\mu\alpha} + \delta g_{\mu\alpha})(\bar{g}^{\alpha\nu} - \delta g^{\alpha\nu}) \\ &= \bar{g}_{\mu\alpha}\bar{g}^{\alpha\nu} - \bar{g}_{\mu\alpha}\delta g^{\alpha\nu} + \delta g_{\mu\alpha}\bar{g}^{\alpha\nu} + O(\delta g^2) \\ &= \delta^\nu_\mu - (\delta g^\nu_\mu - \delta g_\mu^\nu) + O(\delta g^2) \\ &= \delta^\nu_\mu + O(\delta g^2), \end{aligned} \tag{3.12}$$

where in the second to last line we used the fact that perturbed quantities still raise/lower indices with the zeroth order metric - other contributions are higher than first order.

This new perturbed metric can then be propagated through all the other relevant quantities in GR (see Sect. 1.1): Christoffel symbols, Ricci tensor, Einstein tensor, etc. These expressions quickly grow in length and complexity, for example the Christoffel symbols become $\bar{\Gamma}^\lambda_{\mu\nu} + \delta\Gamma^\lambda_{\mu\nu}$ with

$$\delta\Gamma^\lambda_{\mu\nu} = \frac{1}{2}\delta g^{\lambda\rho}\left(\partial_\mu\bar{g}_{\nu\rho} + \partial_\nu\bar{g}_{\rho\mu} - \partial_\rho\bar{g}_{\mu\nu}\right) + \frac{1}{2}\bar{g}^{\lambda\rho}\left(\partial_\mu\delta g_{\nu\rho} + \partial_\nu\delta g_{\rho\mu} - \partial_\rho\delta g_{\mu\nu}\right). \tag{3.13}$$

We leave the remaining steps in the derivation to well established literature, and focus instead on important conceptual simplifications that are important not only for ease of calculation, but also must be used to give meaningful conclusions from the bottom line answers.

3.3.1 SVT Decomposition

> The second, to divide each of the difficulties under examination into as many parts as possible, and as might be necessary for its adequate solution.
>
> –René Descartes, *Discourse on the method of rightly conducting the reason, and seeking truth in the sciences*

> The results obtained so far are repulsively complicated
>
> –Stephen Weinberg, *Cosmology*

Perhaps the most important aspect of CPT for this discussion is that we can consider the decomposition of $\delta g_{\mu\nu}$ into a set of Scalars, Vectors and Tensors (SVT) (Stewart, 1990). More precisely we should say scalars, transverse vectors and transverse traceless tensors, but this will be dropped for ease of writing.

This split can be understood as a generalisation of the Helmholtz decomposition familiar to students of electromagnetism (Griffiths, 2017). There we commonly divide a vector field into a *diverging part*, and a *curling part*:

$$\boldsymbol{E} = \boldsymbol{\nabla}\phi + \boldsymbol{\nabla} \times \boldsymbol{A}. \tag{3.14}$$

If we perform a Fourier transform on this decomposition, then we find

$$\tilde{\boldsymbol{E}} = \boldsymbol{k}\tilde{\phi} + \boldsymbol{k} \times \tilde{\boldsymbol{A}}, \tag{3.15}$$

Together with appropriate boundary conditions these two components fully determine its original vector .

In real space the scalar field contribution can be referred to as the diverging part, since $\boldsymbol{\nabla} \cdot \boldsymbol{\nabla}\phi \neq 0$, while $\boldsymbol{\nabla} \cdot \boldsymbol{\nabla} \times \boldsymbol{A} = 0$. In k-space the diverging field $\tilde{\phi}$ aligns with $\boldsymbol{k}$, while $\tilde{\boldsymbol{A}}$ exists in the orthogonal subspace to $\boldsymbol{k}$, hence receiving the name of *transverse* vector. In heuristic terms, the component of a vector that aligns with $\boldsymbol{k}$ is truly a scalar because no specification of direction was necessary. Note that assuming we are studying electromagnetism in 3 spatial dimensions, we then have one scalar and two vector degrees of freedom.

Extending this concept to GR there is more freedom, since the perturbations take place in a rank 2 tensor. This means that now a tensor *fits* within the perturbations. Again heuristically this is transverse to the $\boldsymbol{k}$ vector and also traceless, since the trace of a matrix can be specified with just a scalar quantity. A full decomposition thus takes the form[2]

$$\delta g_{00} = A, \tag{3.16a}$$

$$\delta g_{0i} = \delta g_{i0} = \partial_i B + E_i^{\mathrm{T}}, \tag{3.16b}$$

$$\delta g_{ij} = C\,\delta_{ij} + \left(\partial_i \partial_j - \frac{1}{3}\delta_{ij}\partial^k \partial_k\right) D + \partial_i F_j^{\mathrm{T}} + \partial_j F_i^{\mathrm{T}} + \Pi_{ij}^{TT}, \tag{3.16c}$$

where A, B, C and D are scalars, E_i^{T} and F_i^{T} are transverse vectors, and Π_{ij}^{TT} is a transverse-traceless tensor. These objects satisfy

$$\partial^i E_i^{\mathrm{T}} = 0, \quad \partial^i F_i^{\mathrm{T}} = 0, \quad \partial^i \Pi_{ij}^{TT} = \partial^j \Pi_{ij}^{TT} = 0, \quad \Pi_{ij}^{TT} = \Pi_{ji}^{TT}. \tag{3.17}$$

This decomposition still contains all the information of the metric, and is therefore explicitly symmetric. A further sanity check is to count the unconstrained components:

- One component from each scalar, four in total.
- Two components from each vector, since one orthogonality constraint is applied. Again four in total
- Two components from the symmetric tensor, since this has orthogonality constraints in each row (-3) and in being traceless (-1).

This totals ten, as expected from a symmetric four dimensional rank two tensor $\delta g_{\mu\nu}$. This means we are free to use this decomposition in place of the full tensor whenever this is convenient.

This innocuous looking decomposition turns out to be extremely powerful, given the way that first order perturbations couple to one another. Roughly speaking, and specific to first order, perturbations of a given type only source their own type. For example, Π type perturbations are only sourced by anisotropic energy distributions.

---

[2] Note that this decomposition has many different conventions in the literature. For example, sometimes the spatial Laplacian is not subtracted from D, making it a contribution to the trace of δg_{ij}. Other times $C \to \frac{C}{3}$, making it exactly equal to the trace of the spatial tensor.

This is exemplified by the origin of gravitational waves in orbiting binary systems (LIGO Scientific Collaboration & Virgo Collaboration, 2016). Consider also that spherically symmetric bodies are prevented from creating gravitational waves as a result of Birkhoff's Theorem (Birkhoff & Langer, 1923). Similarly, perturbations to pressure and energy are associated with the scalar sector, and will be our prime focus moving forward.

3.3.2 Gauge Choice

> Everything we hear is an opinion, not a fact. Everything we see is a perspective, not the truth.
>
> –Marcus Aurelius, *Meditations*

One of the subtleties that remains to be discussed is that of choosing a gauge (Mukhanov et al., 1992; Ma & Bertschinger, 1995). This is related with the discussion in Sect. 1.1, where the concept of coordinate invariance is fundamental to the formalism of General Relativity. The concept of a gauge is essentially the process of choosing a coordinate frame in which the perturbations are being studied, while ensuring that any bottom line results are independent of this choice. As an example, imagine labouring over a perturbed ΛCDM Universe with a small velocity field in a single spatial direction. It is intuitive that this solution could have been achieved simply with a change of coordinates on the background – a small Lorentz boost.

Now let us express this idea mathematically. We need to see how the components of a general tensor, T, will change under a small (perturbative) change of coordinates. The bottom line is that these transformations are given by Lie derivatives with respect to a vector field, ξ, which is assumed to be small in its components:

$$\begin{aligned} T &\to T + (\Delta T)_\xi, \\ (\Delta T)_\xi &= \mathcal{L}_\xi T, \end{aligned} \tag{3.18}$$

where the Lie derivative $\mathcal{L}_\xi$ of an (m, n) tensor is given by

$$\begin{aligned}(\mathcal{L}_\xi T)^{a_1\ldots a_m}{}_{b_1\ldots b_n} &= \xi^c(\partial_c T^{a_1\ldots a_m}{}_{b_1\ldots b_n}) \\ &\quad - (\partial_c \xi^{a_1})T^{c\ldots a_m}{}_{b_1\ldots b_n} - \ldots - (\partial_c \xi^{a_m})T^{a_1\ldots c}{}_{b_1\ldots b_n} \\ &\quad + (\partial_{b_1}\xi^c)T^{a_1\ldots a_m}{}_{c\ldots b_n} + \ldots + (\partial_{b_n}\xi^c)T^{a_1\ldots a_m}{}_{b_1\ldots c}.\end{aligned} \tag{3.19}$$

This arises due to a correspondence between diffeomorphisms and vector fields, a full discussion of which can be found in appendix B of Carroll (2004). In essence, since the change of coordinates is necessarily smooth, and small in magnitude, this is equivalent to moving objects along flow lines given by some smooth vector field ξ, which also has perturbatively small components. Following the convention used for the metric, let's decompose ξ into 1 + 3 components[3] $\xi^\mu \to (T, L^i)$, and evaluate the change of the metric, noting that $\mathcal{L}_\xi \delta g_{\mu\nu} = 0$:

$$\begin{aligned}(\Delta g_{\mu\nu})_\xi = \mathcal{L}_\xi g_{\mu\nu} = \mathcal{L}_\xi \bar{g}_{\mu\nu} &= \xi^\alpha\left(\partial_\alpha \bar{g}_{\mu\nu}\right) + \left(\partial_\mu \xi^\alpha\right)\bar{g}_{\alpha\nu} + \left(\partial_\nu \xi^\alpha\right)\bar{g}_{\mu\alpha} \\ &= T\dot{\bar{g}}_{\mu\nu} + \left(\partial_\mu T\right)\bar{g}_{0\nu} + \left(\partial_\nu T\right)\bar{g}_{\mu 0}\left(\partial_\mu L^k\right)\bar{g}_{k\nu} + \left(\partial_\nu L^k\right)\bar{g}_{\mu k},\end{aligned} \tag{3.20}$$

where we set $L\left(\partial_k \bar{g}_{\mu\nu}\right) = 0$, since the FLRW metric has no spatial dependence due to the flatness. Evaluating this in time/space components of the metric, and further decomposing $L_i = \partial_i L + L_i^{\mathrm{T}}$ (see Sect. 3.3.1) we get

$$(\Delta g_{00})_\xi = 2\dot{T}, \tag{3.21a}$$

$$(\Delta g_{i0})_\xi = (\Delta g_{0i})_\xi = \left(\partial_t - 2\frac{\dot{a}}{a}\right)\left(\partial_i L + L_i^{\mathrm{T}}\right) - \partial_i T, \tag{3.21b}$$

$$(\Delta g_{ij})_\xi = 2a\dot{a}T\delta_{ij} + \partial_i L_j^{\mathrm{T}} + \partial_j L_i^{\mathrm{T}} + 2\partial_i\partial_j L. \tag{3.21c}$$

Applying these gauge transformations to the components of the SVT split $\delta g_{\mu\nu}$ via

---

[3] See Gourgoulhon (2012) and references therein for a clear view of the 3 + 1 formalism in more detail and rigour

Eq. (3.18), we find:

$$\begin{aligned}
\left(\Delta_\xi g_{00}\right):&\quad A \to A + 2\dot{T}, && (3.22\text{a})\\
\left(\Delta_\xi g_{0i}\right):&\quad B \to B + \dot{L} - 2\frac{\dot{a}}{a}L - T, && (3.22\text{b})\\
&\quad E_i^{\mathrm{T}} \to E_i^{\mathrm{T}} + \dot{L}_i^{\mathrm{T}} - 2\frac{\dot{a}}{a}L_i^{\mathrm{T}}, && (3.22\text{c})\\
\left(\Delta_\xi g_{ij}\right):&\quad C \to C + 2a\dot{a}T + \frac{1}{3}\partial^k\partial_k L, && (3.22\text{d})\\
&\quad D \to D + L, && (3.22\text{e})\\
&\quad F_i^{\mathrm{T}} \to F_i^{\mathrm{T}} + L_i^{\mathrm{T}}, && (3.22\text{f})\\
&\quad \Pi_{ij}^{TT} \to \Pi_{ij}^{TT}. && (3.22\text{g})
\end{aligned}$$

It is notable that Π_{ij}^{TT} can't be changed through a gauge transformation, meaning that tensor perturbations (i.e. gravitational waves) are not frame-dependent objects (Weinberg, 2008).

The bottom line is that our perturbed tensors will change depending on the reference frame chosen. On the one hand, this is a problem, giving us extra work in checking any derived results are not just coordinate artefacts. On the other hand, it gives some freedom, as we can choose the 4 degrees of freedom (1 in T, 3 in L_i). Moving forward, two general philosophies can be adopted: combine the perturbations in a way that gauge dependence cancels, or specify which gauge you are in by fixing some components.

For the vector perturbations, it should be easy to see that defining

$$\begin{gathered}
\mathcal{V}_i^{\mathrm{T}} = E_i^{\mathrm{T}} - \dot{F}_i^{\mathrm{T}} + 2\frac{\dot{a}}{a}F_i^{\mathrm{T}},\\
\Delta_\xi \mathcal{V}_i = 0,
\end{gathered} \qquad (3.23)$$

gives a gauge invariant vector perturbation.

Similar gauge invariant quantities exist for scalar perturbations but are more complicated (Bardeen, 1980). Thus we will apply the second of the aforementioned approaches, where we simply choose a gauge and understand the corresponding artefacts in the bottom line answers. Two common gauges exist: the synchronous gauge, and the conformal Newtonian gauge (Ma & Bertschinger, 1995). The latter will be introduced here and used throughout the thesis. In this choice of gauge we choose L such that $D = 0$, and T such that $B = 0$. This leaves A and C, which we will rename to more easily match other common notations:

$$A = -2\Psi, \quad C = -2a^2\Phi. \tag{3.24}$$

It is common in studying the CMB to focus only on scalar potentials, ignoring vector and tensor perturbations. This is often sufficient since photons couple strongly to scalar perturbations through their overdensities, while properties like polarisation are influenced by tensors. We will thus often make use of the metric (in conformal Newtonian gauge)

$$\mathrm{d}s^2 = -(1 + 2\Psi)\mathrm{d}t^2 + a^2(t)\,(1 + 2\Phi)\,\mathrm{d}\boldsymbol{x}^2. \tag{3.25}$$

3.3.3 Stress Energy Tensor

So far we have perturbed the metric, performed the SVT split and selected a gauge for these perturbations. This process relates specifically to the LHS of Einstein's field equation (Eq. 1.11), and we have yet to discuss the corresponding perturbations to the stress energy tensor. Luckily, much of the previous discussion will be useful here. Firstly, we will only discuss the 4 scalar degrees of freedom, which here correspond to relatively intuitive quantities. Secondly, the choice of gauge has been used to cancel geometric quantities. This means the similar gauge terms on the RHS have already been chosen, and will only serve to rescale the quantities with no full cancellations. To be specific, local energy over/underdensities $\delta\rho$ will be slightly different in the synchronous and conformal Newtonian gauge, but correspond to the same physical entity, and in neither case will be set to 0.

Discussing the RHS equivalents of A, B, C and D in order we have $\delta\rho$, $\delta\theta$, δp and

$\delta\sigma$. The first quantity is clear, a perturbation to the energy content in some patch of the Universe. The second quantity is related to a velocity perturbation. This is explicitly perturbative since in background ΛCDM no velocities can exist, as this would violate local isotropy. The third quantity is on the trace of T_{ij}, making it the perturbation to the isotropic pressure. The last quantity is the scalar quantity present in T_{ij}, but this time off axis, corresponding to some anisotropic stress. This last component is generally not sourced by matter, but can be sourced by radiation and neutrino fluids (when they cease to behave as a perfect fluid, see Sect 1.3). Note that all quantities here are written with δX, despite X not always corresponding to perturbed versions of a background quantity, which differs in notation from Ma & Bertschinger (1995) where θ is implicitly small.

3.3.4 Bottom line

Carrying out all the above techniques yields four equations governing the evolution of scalar gravitational potentials, although only two are independent. Following Ma & Bertschinger (1995) but with the variable convention[4] of chapter 7 we have

$$-k^2\Phi + 3\frac{a'}{a}\left(-\Phi' + \frac{a'}{a}\Psi\right) = \sum_\lambda a^2 16\pi G\delta\rho_\lambda, \tag{3.26a}$$

$$k\left(-\Phi' + \frac{a'}{a}\Psi\right) = \sum_\lambda a^2 12\pi G(\bar{\rho} + \bar{P})\delta\theta_\lambda, \tag{3.26b}$$

$$-\Phi'' + \frac{a'}{a}\left(\Psi' - 2\Phi'\right) + \left(2\frac{a''}{a} - \frac{a'^2}{a^2}\right)\Psi - \frac{k^2}{3}\left(\Phi + \Psi\right) = \sum_\lambda a^2\frac{4\pi G}{3}\delta p_\lambda, \tag{3.26c}$$

$$-k^2\left(\Phi + \Psi\right) = \sum_\lambda a^2 24\pi G\left(\bar{\rho} + \bar{P}\right)\delta\sigma_\lambda, \tag{3.26d}$$

where we have explicitly defined the stress-energy perturbations as the sum over relevant quantities from different sectors λ (i.e. photons, neutrinos, dark matter, baryons). For example, writing the quadrupole of neutrinos as $\mathcal{N}_2^{(1)}$ we write the shear stress as $\sum_\lambda \delta\sigma_\lambda = \Theta_2^{(1)} + \mathcal{N}_2^{(1)}$, considering that only relativistic particles carry shear stress.

---

[4] Compared to Ma & Bertschinger (1995) we have $\Phi \to -\Phi$. Furthermore we note that for relativistic particles we have $\delta_\mathrm{r} \to 4\delta\rho_\mathrm{r}$, $\theta_\mathrm{r} \to 3k\delta\theta_\mathrm{r}$, $\sigma_\mathrm{r} \to 2\delta\sigma_\mathrm{r}$.

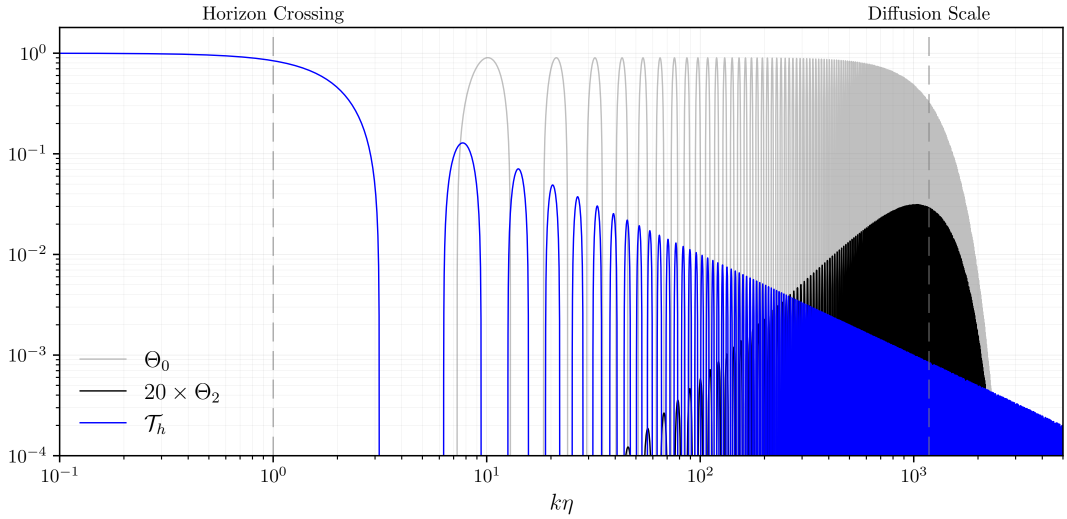


Figure 3.2: A figure showing how scalar and tensor perturbations differ in their dynamics. Θ_0 *shows a constant amplitude envelope within which is oscillates due to a standing wave behaviour with electrons. The ending of that wave is characterised by a growing quadrupole moment, which is a useful quantity for understanding Silk damping and blackbody mixing (Chluba et al., 2012; Silk, 1968). The tensor perturbation on the the other hand decays immediately after horizon crossing but at a slower rate. This causes a stretching of the relevant tensor window for SD constraints compared to scalar perturbations (Chluba et al, 2019; Chluba et al., 2015a).*

3.4 Evolution of primordial perturbations

The results of Sect. 3.2 and Sect. 3.3 together give the full picture of primordial photon perturbations - at least as far as temperature anisotropies is concerned. A full solution of these equations however is not easily accessible to analytic treatments (Hu & Sugiyama, 1995, 1996), and often numerical solutions are employed (Lewis et al., 2000; Lesgourgues, 2011). For the purposes of this thesis it is most useful to focus on limiting cases and derive heuristic and conceptual understanding of these primordial waves. We will do so without laborious derivations, and with more of a focus on qualitative results.

3.4.1 Horizon crossing

The concept of horizon crossing can classically be thought of in terms of causality. Essentially perturbations with a wavelength larger than the current Hubble horizon have not had enough time for light signals to travel from one peak to another. This leads to a situation where these waves are frozen and static till the horizon has expanded. In fact there is the alternative solution of a power law decay, but the important aspect here is the independence of wavelength (a field can "everywhere locally" decay without violating any causality condition).

Mathematically this is best seen in the evolution of tensorial perturbations[5] which in the absence of any anisotropic stress evolve according to (see Weinberg, 2004; Dicus & Repko, 2005)

$$h''(\eta) + 2\frac{a'}{a}h'(\eta) + k^2 h(\eta) \approx 0. \tag{3.27}$$

Furthermore we will pursue solutions in the radiation (really, relativistic) dominated era with $a = \eta H_0 \sqrt{\Omega_{\rm rel}}$

$$h''(\eta) + \frac{2}{\eta}h'(\eta) + k^2 h(\eta) \approx 0. \tag{3.28}$$

The second term's coefficient is the Hubble rate, which falls with time, while the third term is simply multiplied by a constant which we know to be associated with wave like behaviour. It is clear that the second term dominates the early time, while the final term provides late time evolution. We can thus pursue solutions initially to

$$h''(\eta) + \frac{2}{\eta}h'(\eta) \approx 0, \tag{3.29}$$

which gives $h(\eta) = A - B/\eta$, with A and B being arbitrary constants. This behaviour is quite characteristic, with one constant mode and one decaying mode at early time. Note that the second solution has a singularity for $\eta \to 0$, however it could in principle be sourced in an unproblematic way at some finite η.

Returning to the solution to Eq. 3.28 we can find $h(\eta) = A j_0(k\eta) - B y_0(k\eta)$ (Watanabe & Komatsu, 2006; Kite et al., 2022), where j_0 and y_0 are the spherical Bessel functions of the first and second kind respectively and A, B are again arbitrary constants. These solutions have similar properties to the solutions above, with the y_0 function have an early time divergence and decaying behaviour. Now however the j_0 mode enters an oscillatory phase after horizon crossing $k\eta \sim 1$ which is characterised by the third term in the ODE.

---

[5] Tensorial solutions are greatly expanded upon in chapter 5, Kite et al. (2022).

Scalar perturbations follow a similar mathematics, with a constant superhorizon mode and a subsequent wave behaviour following the time of crossing. Photon temperature perturbations (scalars) and gravitational waves (tensors) are shown in Fig. 3.2, where the moment of horizon crossing is indicated, showing common behaviour between these two conceptually very distinct waves.

This moment of horizon crossing is important for the CMB anisotropies, where large patches of the sky are typically dominated by modes which crossed the horizon more recently. Even for tensors there is a distinct bend in today's tensor energy spectrum associated with modes crossing horizon before or after the transition between a radiation and matter dominated Universe (Watanabe & Komatsu, 2006; Saikawa & Shirai, 2018; Kite et al., 2022).

3.4.2 Tight coupling

Once the perturbations have crossed horizon they enter an oscillatory phase. Here we now see big differences between the tensors and scalars, as can be appreciated in Fig. 3.2. This difference is because the photons and baryons form a stable oscillation, with gravity and electromagnetic repulsion forming the restoring forces. We can understand this in terms of a single *tightly coupled* fluid, where the two components behave as one. Eventually this marriage of photons and charged baryons fails, and there is a decay of the mode. Tensor waves have no such supporting sources, and simply decay with a power law, which is slower than the exponential decay in the temperature perturbations once they do decay.

In the tight coupling limit we have $\tau' \gg k$, implying that Thomson scattering is very effective at isotropising the medium, and keep the oscillations in the monopole and dipole of the relevant fluids. An example of how there is effectively one fluid is seen in

Eq. 3.8, where the term proportional to τ' drives $\Theta_1 \approx \beta/3$, implying the photons and baryons share a common velocity.

The photon hierarchy in this limit gives (Dodelson, 2003)

$$\begin{aligned} \Theta_0'' + \frac{a'}{a}\frac{R}{1+R}\Theta_0' + k^2 c_s^2 \Theta_0 &= F(k,\eta), \\ F(k,\eta) &= -\frac{k^2}{3}\Psi - \frac{a'}{a}\frac{R}{1+R}\Phi - \Phi''. \end{aligned} \tag{3.30}$$

We will not analyse this equation in too much detail, but will limit ourselves to pointing out its structure as a forced and damped harmonic oscillator. The forcing in this case being a time dependent combination of gravitational potentials. Here $R = 3\rho_b/4\rho_\gamma$ gives the ratio of baryon and photon energies, thus giving the damping term a physical interpretation known as baryon loading. Essentially the greater this ratio is the more mass the photon-baryon plasma has, damping the oscillations.

Provided this forced oscillator solution requires a highly isotropised photon fluid (e.g. no quadrupole moment), it will break down when $\tau' \gg k$ ceases to be true.

3.4.3 Diffusion scale

The derivation of Eq. 3.30 required establishing the hierarchy of photon moments, truncating at the dipole, and then eliminating that dipole from the equation set. To understand the breakdown of the tight coupling limit we much return to the hierarchy and include the quadrupole term also. This prevents us writing down a single equation for the monopole and forces us into the full coupled system. Again leaving the details to literature (see e.g. chapter 9 of Dodelson, 2003) we have

$$\hat{\Theta}_0' + k\hat{\Theta}_1 = 0, \tag{3.31}$$

$$\hat{\Theta}_1' + k\left(\frac{2}{3}\hat{\Theta}_2 - \frac{1}{3}\hat{\Theta}_0\right) = \tau'\left(\hat{\Theta}_1 - \frac{\hat{\beta}}{3}\right), \tag{3.32}$$

$$\hat{\Theta}_2' - \frac{2k}{5}\hat{\Theta}_1 = \frac{9}{10}\tau'\hat{\Theta}_2. \tag{3.33}$$

Here we have not neglected Θ_2 as before. Now Θ_3 is the first neglected term, being a factor of $\sim k/\tau'$ smaller than Θ_2. Assuming we're still close to the tight coupling era, we can still use $\tau' > k$ which gives a quasi-stationary evolution $\Theta_2' \approx 0 \implies \Theta_2 \approx -4k/9\tau'\,\Theta_1$. This expressions makes it so the quadrupole is a sort of *leaking of energy* form the hierarchy, leading to an eventual exponential decay of the temperature perturbations. Concretely we can find

$$\begin{aligned} \Theta_0,\, \Theta_1 &\sim \exp\left[ik \int \mathrm{d}\tilde{\eta}\, c_s(\tilde{\eta})\right] \exp\left(-\frac{k^2}{k_{\mathrm{D}}^2}\right), \\ k_{\mathrm{D}}^{-2} &= \int_0^{\eta} \frac{\mathrm{d}\tilde{\eta}}{6(1+R)\tau'(\tilde{\eta})} \left[\frac{R^2}{1+R} + \frac{8}{9}\right], \\ c_s(\tilde{\eta}) &= \sqrt{\frac{1}{3(1+R(\tilde{\eta}))}}, \end{aligned} \tag{3.34}$$

where the details of the 8/9 factor is associated with the inclusion of polarization, not discussed here. The integral governing the oscillation speed includes the sound speed in the fluid c_s. k_{D} provides the diffusion scale drawn in Fig. 3.2. The diffusion scale shrinks with time, and when it evolves enough so that the scale of a given mode k surpasses it, we see an exponential decay of the perturbations. This decay correlates with a rapid excitation of the quadrupole moment.

This whole process has a visual classic analogue. While the photon fluid is locked in oscillation with baryons in a macroscopic sense, there is a random walk being undertaken by the photons at a microscopic level. The mean free path of this walk can be found to follow $\lambda_{\mathrm{MFP}} \sim 1/\tau'$, which provides a nice physical interpretation to the limit $k/\tau' = k\lambda_{\mathrm{MFP}} << 1$. Basically these Fourier modes are concerned with lengths over which photons have not been able to diffuse The diffusion process heuristically implies a mixing of two patches with different temperatures, and therefore erases the differences (see Fig. 2.4). From our discussion in chapter 2 we know this implies the creation of a y distortion. This was studied in detail in Chluba et al. (2012) yielding an energy injection of (see also Chluba et al., 2015a, for succinct summary)

$$\frac{\mathrm{d}Q}{\mathrm{d}\tau}(\eta) \approx 4 \int \frac{k^2 \mathrm{d}k}{2\pi^2} P(k) \frac{9}{2} \Theta_2(k,\eta)^2, \tag{3.35}$$

where the power spectrum $P(k)$ will be formally introduced in Sect 3.6.

3.4.4 Free-streaming

Primordial perturbations in the photon fluid are not necessarily condemned to self erasure though photon diffusion. There is an alternative fate for patches which survive long enough for the primordial plasma to dissipate, cool, and undergo recombination. With electrons and protons combining to make neutral hydrogen the photons can then free-stream, and the structure of the standing acoustic waves is locked in. In this sense there is a "snapshot" of the primordial plasma which we today call the CMB. It is typically the larger scale modes (small k) that see this fate, with smaller scale (larger k) dissipating too early. Note however that the smallest of k have not yet crossed horizon, providing a strong cutoff for modes approaching $k \sim 1/\eta_0$.

This process of free-streaming can be seen in the photon hierarchy Eq. 3.6 by simply neglecting any terms with τ'. This leaves the hierarchy of multiples coupled with k. This leads to a heuristic picture where energy tends to flow up to higher multipoles (a plane wave can be represented as a summation of Legendre polynomials), but Thomson scattering provided an isotropising effect which ceases at recombination. This is why when we measure angular patches on the sky at modern times we see $\ell \sim 1000$, despite the fact that in the primordial plasma anything higher than the quadrupole is seldom important.

It turns out that once free-streaming starts it is no longer computationally (or analytically) efficient to understand the equation set as many thousands of coupled differential equations, and there are alternative approaches that regain the view of plane waves propagating. In particular there is a line of sight integral which drastically simplifies and optimises codes modelling the Boltzmann hierarchy (Seljak & Zaldarriaga, 1996).

3.5 Integral Equation Approach

We briefly mention that although the multipole hierarchy is the most common approach to solving early universe perturbations (e.g. Lesgourgues, 2011; Lewis et al., 2000), occasionally a formulation in terms of integral equations shows some benefits. An early use of this approach was shown in Weinberg (2004), which matches the approach used in chapter 5 for studying the damping of tensor perturbations by neutrinos. The integral equation approach for scalar perturbations was introduced in Weinberg (2006), but generalised and numerically benchmarked in recent work by Kamionkowski (2021). This work showed that there is generally an equivalence between the infinite multipole hierarchies, that have to be truncated for numerical implementation, and the integral equation approach, which can be solved without truncation via an iterative numerical scheme. While the iterative approach developed in Kamionkowski (2021) is not equal in all details to that shown in chapter 5, we make use of the same principles.

In the case of photons it is known that to get the correct spectrum of CMB anisotropies within the Boltzmann hierarchy approach it is sufficient to use $\ell_{\rm max} \sim 30$, a surprising fact considering that we observe patches in the sky with $\ell \gg 1000$. This comes down to the fact that most information comes from a visibility function which peaks around recombination (see Seljak & Zaldarriaga, 1996; Callin, 2006, and discussion of numerical implementation within chapter 8). Getting the multipole hierarchy correct within that visibility window is the most important part of the calculation, allowing one to truncate the hierarchy at low ℓ, even if at later times the hierarchy would naturally grow to $\ell > 1000$. The trade-off between the multipole hierarchy and the integral equation approaches is the solution of ~ 30 extra equations for each hierarchy or a few iterations involving integrals. The claim within Kamionkowski (2021) is that the latter choice could accelerate Boltzmann solvers.

Similar arguments *would* apply for studying the multiples of neutrino perturbations in the sky (a futile endeavour in itself from an observational viewpoint). Performing such a calculation would involve defining similar visibility functions as for photons, and connecting to a line-of-sight integration approach for the free-streaming neutrinos. This however isn't the goal in chapter 5. Instead we focus on the energy lost from the gravitational waves. To do this within the Boltzmann hierarchy we would have to start our calculations in a phase where neutrinos are not free-streaming (in order to define the initial conditions) and then solve coupled equations potentially to very late times (e.g. till recombination). In this case we would most likely require $\ell_{\max} \gg 30$ to achieve accurate results, and thus the integral equation approach is simpler both conceptually and numerically.

3.6 The Primordial Power Spectrum

Knowing how the perturbations evolve is only half the story, since the initial conditions are still to be defined. This problem will be undertaken in a statistical manner by discussing the moments of a probability distribution function. We define the ensemble average, $\langle ... \rangle$, as the average over an ensemble of hypothetical Universes, and by the ergodic hypothesis equate this to an average over space in our Universe. This however is more easily evaluated using Fourier transforms, which is equivalent through Parseval's theorem (Martínez et al., 2009).

It may come as no surprise that all metric perturbation components have $\langle \delta g_{\mu\nu} \rangle = \langle \delta T_{\mu\nu} \rangle = 0$. To quickly justify this however, consider that any overall shift to a scalar component like Φ would just be subtracted, and added to the background energy instead. Vector perturbations, on the other hand, select directions in space and therefore can't be subtracted as simply. However, it can be argued that on average no direction should be picked over any other, and therefore $\langle \mathcal{V}_i^{\mathrm{T}} \rangle = 0$. Similar arguments hold for tensor perturbations.

We can rewrite a generic Fourier transformed perturbation field, $X(\boldsymbol{k}, \eta)$, as the product of a deterministic transfer function, $\mathcal{T}_X(\boldsymbol{k}, \eta)$, and some initial condition, $\zeta(\boldsymbol{k})$, which is probabilistic in nature. Thus at linear order we have $X(\boldsymbol{k}, t) = \zeta(\boldsymbol{k}) \mathcal{T}_X(\boldsymbol{k}, t)$ (See Pettinari, 2016, for discussion of second order transfer functions). To avoid a cumbersome set of transfer functions it is common to simply use the same variable name (i.e. $X(\boldsymbol{k}, t) \to \zeta(\boldsymbol{k})X(\boldsymbol{k}, t)$). In this language the above discussion then amounts to stating the first moment of $\zeta(\boldsymbol{k})$, $\langle\zeta(\boldsymbol{k})\rangle$ vanishes. The second moment of the perturbations can be non-zero however, and are given as a primordial power spectrum via $\langle\zeta^*(\boldsymbol{k})\zeta(\boldsymbol{k}')\rangle = (2\pi)^3\,\delta^{(3)}(\boldsymbol{k} - \boldsymbol{k}')P(k)$.

The primordial power spectrum essentially encodes the statistics of the initial condition of fluctuations (total amplitude and wavelength dependence) for some component of the cosmic inventory. Once a power spectrum has dictated the initial seeds for one component of the perturbation picture we define other variables in the relative proportions (See e.g. Chluba & Grin, 2013, for discussions of adiabatic and isocurvature initial conditions and relevant spectral distortion constraints).

Higher moments of $\zeta(\boldsymbol{k})$ are all consistent with 0 in currently available experimental data, making what is known as a Gaussian power spectrum. There is, however, a lot of discussion in the literature of non-gaussianities in the CMB spectrum. If discovered these would probe very early physics, and would shed light on the inflation, the currently favoured theory of the earliest moments of the Universe (Celoria & Matarrese, 2018).

The primordial scalar power spectrum as currently understood[6] is usually defined around a pivot scale k_0, and is written such that it is flat except for a small tilt, dictated by a spectral index $n_\mathrm{s} \approx 1$, and some small curvature, given by a running of the spectral index $n_\mathrm{run} \approx 0$ (Planck Collaboration, 2019, 2018b):

$$P(k) = \frac{2\pi^2}{k^3} A \left(\frac{k}{k_0}\right)^{n_\mathrm{s}-1+\frac{1}{2}n_\mathrm{run}\ln\left(\frac{k}{k_0}\right)}. \tag{3.36}$$

---

[6] Hypothetical models will often have spikes at certain scales, but the currently observed standard physics is limited to a simple straight line for scalar perturbations. Currently the shape of the tensor power spectrum is unconstrained, with only upper bounds on tensor to scalar ratios (Tristram et al., 2022).

Preface to Published Work

The following five chapters each contain one paper - the first two published, and the final three in various stages of the peer review process. These papers all relate back to the theory introduced in the previous introductory chapters. While before each paper we briefly summarise the content, it is useful to anticipate how the papers connect to one another as well as the previous chapters. Both in terms of chronology and content there are in essence two groups: the first two papers and the final three.

The first two papers consist of an exploration of the primordial tensorial perturbations which permeate the Universe, and their potential signal in the form of spectral distortions. The perturbations themselves have not appeared very much throughout chapter 3, since for reasons of cosmological interest and pedagogy we stuck mostly to scalar perturbations. We can point out however that these waves are simply the tensorial component of the spatial metric discussed in Sect. 3.3.1. We briefly used these perturbations as en example to study horizon crossing in Sect. 3.4.1, and one numerically calculated tensor transfer function was shown in Fig. 3.2. Readers interested in the primordial tensor waves will hopefully find the second paper (chapter. 5) to be a clear and somewhat complete introduction. The motivation for that paper however, in connection to this thesis, is best found in the first paper (chapter. 4), where we study the small albeit non-zero energy injection to the photon bath originating from tensor modes. This occurs through a sourcing of an anisotropic stress in the photon fluid, and can lead to the formation of an observable μ-distortion signal for some exotic physical scenarios. The first paper thus unites concepts from both chapter 2 and chapter 3. In summary, these two papers study standard cosmological perturbations and how they can mix and source SD signals.

The final three papers unite SD and CPT in a completely different and more profound way. The Boltzmann hierarchy at the start of chapter 3 initially contained a momentum or frequency dependence (see Eq.(3.5)). We promptly simplified this to local inhomogeneities in the temperature of photons, thus eliminating the frequency space (see Eq.(3.6)). This move was subtle, and essentially neglected what we had discussed throughout chapter 2. The goal of the final three papers is to do away with that *collapsing* of frequency space in studying cosmological perturbations, and allow for a general frequency evolution of the spatially varying photon gas. The first paper in that series (chapter 6) has nothing to do with the spatial part of the overall goal, and instead focusses on simply building a basis of spectral shapes. These shapes are, however, chosen in such a way that the spatial picture comes much easier, while still replicating the expected homogeneous results (the three era picture from Sect 2.3). The second paper (chapter 7) can then take that spectral basis and pass it through the machinery of CPT discussed throughout chapter 3. The final equations are a generalisation of Eq. (3.6), allowing a full spectro-spatial evolution of the CMB sky. The final paper (chapter 8) contains the numerical results derived from the generalised Boltzmann hierarchy. We distil the physics and pick apart aspects of the transfer function evolution (akin to what we saw in Fig. 3.2) and study cross correlations of temperature anisotropies with distortion anisotropies. The bottom line of that paper is that observing anisotropies reveals something about the average photon spectrum across the sky, thus proving a way to improve upon the *COBE/FIRAS* results with future instruments resembling (but more advanced than) *Planck*.

4

Paper I: Bridging the gap: spectral distortions meet gravitational waves

This paper revisits the calculations of Ota et al. (2014) and Chluba et al. (2015a) to not only update the results, but to also cast the language used into that which is more familiar in the newly emerging gravitational wave literature.

In essence the paper discusses the possibility of primordial tensor perturbations (see chapter 3) dissipating energy into the primordial plasma and thus creating primordial spectral distortions (see chapter 2). The physics involved in this process can be packaged up into a pretabulated window function, which is simply integrated with a primordial tensor power spectrum to derive a μ-distortion amplitude. This method however is insufficient for tensor creation mechanisms which are not primordial, even if they occur very early. We argue that this can be modelled with a power spectrum still, just with an appropriate step function and redshifting included. We pretabulate the window function for various finite redshift tensor injections and provide a `Python` tool which can interpolate on the results.

With these window functions in place we show that the SD constraints bridge a gap between large-scale Cosmological constraints from CMB anisotropies and small-scale astrophysical constraints from pulsar timing arrays.

To further demonstrate the power of SD constraints on tensor backgrounds we undertake some case studies, including cosmic string networks, axion models tuned to produce tensorial decays, and phase transitions.

Mon. Not. R. Astron. Soc. **000**, 000–000 (0000) Printed 30 July 2021 (MN LaTeX style file v2.2)

Bridging the gap: spectral distortions meet gravitational waves

Thomas Kite[1]⋆, Andrea Ravenni[1]†, Subodh P. Patil[2]‡ and Jens Chluba[1]§
[1] *Jodrell Bank Centre for Astrophysics, School of Physics and Astronomy, The University of Manchester, Manchester, M13 9PL, U.K.*
[2] *Instituut-Lorentz for Theoretical Physics, Leiden University, 2333 CA Leiden, The Netherlands*

arXiv:2010.00040v2 [astro-ph.CO] 29 Jul 2021

Accepted 2020 –. Received 2020 –

ABSTRACT
Gravitational waves (GWs) have the potential to probe the entirety of cosmological history due to their nearly perfect decoupling from the thermal bath and any intervening matter after emission. In recent years, GW cosmology has evolved from merely being an exciting prospect to an actively pursued avenue for discovery, and the early results are very promising. As we highlight in this paper, spectral distortions (SDs) of the cosmic microwave background (CMB) uniquely probe GWs over six decades in frequency, bridging the gap between astrophysical high- and cosmological low-frequency measurements. This means SDs will not only complement other GW observations, but will be the sole probe of physical processes at certain scales. To illustrate this point, we explore the constraining power of various proposed SD missions on a number of phenomenological scenarios: early-universe phase transitions (PTs), GW production via the dynamics of SU(2) and ultra-light U(1) axions, and cosmic string (CS) network collapse. We highlight how some regions of parameter space were already excluded with data from *COBE/FIRAS*, taken over two decades ago. To facilitate the implementation of SD constraints in arbitrary models we provide `GW2SD`. This tool calculates the window function, which easily maps a GW spectrum to a SD amplitude, thus opening another portal for GW cosmology with SDs, with wide reaching implications for particle physics phenomenology.

Key words: cosmology: theory — gravitational waves — early Universe — inflation — cosmic background radiation.

1 INTRODUCTION

Gravitational wave (GW) astronomy has become a reality. The now routine detection of compact object mergers by the LIGO/Virgo collaboration (Abbott et al. 2019) has made, for good reasons, the study of GWs one of the most active and current topics in cosmology and astrophysics. Ongoing and planned observations of the tensor perturbation power spectrum currently span some 21 orders of magnitudes of frequency: From cosmic microwave background (CMB) upper limits on primordial B-modes (Ade et al. 2018; Aghanim et al. 2020) measurements at the lowest frequencies, to interferometry detections of GWs (e.g., Abbott et al. 2020b,a) and Pulsar Timing Array (PTA) measurements (e.g., Perera et al. 2019; Alam et al. 2020) at higher frequencies. In the next few years, a plethora of experiments will test different scales between these extremes (e.g., Campeti et al. 2021, for overview).

Many physical processes can indeed lead to detectable tensor perturbations (see Caprini & Figueroa 2018, for review). These include GWs from phase transitions (Caprini & Figueroa 2018; Nakai et al. 2020), early universe gauge field production (Dimastrogiovanni et al. 2017; Machado et al. 2019; Machado et al. 2020), and cosmic string networks (Buchmuller et al. 2019). Given these exciting theoretical developments, it is interesting to ask which cosmological and astrophysical probes can help constrain these different scenarios. In this paper, we show that CMB spectral distortions (SDs) can provide complementary information at frequencies $f = 10^{-15}$–10^{-9} Hz unavailable to other probes. In this way, SDs offer a bridge between scales probed by next-generation CMB surveys (e.g., Ade et al. 2019; Hazumi et al. 2019; Delabrouille et al. 2019), and astrophysical GW observatories such as current (e.g. Perera et al. 2019) and future (e.g. Weltman et al. 2020) PTA measurements.

How do CMB SDs constrain tensor perturbations at the scales that they do? Spectral distortions are created by mechanisms that lead to energy release into the photon-baryon fluid at redshifts $z \lesssim 2 \times 10^6$, when thermalization processes cease to be efficient (Zeldovich & Sunyaev 1969; Sunyaev & Zeldovich 1970; Illarionov & Sunyaev 1975; Danese & de Zotti 1982; Burigana et al. 1991; Hu & Silk 1993; Chluba & Sunyaev 2012). Many sources of distortions exist within standard ΛCDM cosmology as well as scenarios invoking new physics (see Chluba et al. 2019b, for broad overview), and innovative experimental concepts (Kogut et al. 2016, 2019; Chluba et al. 2019a) have now reached critical thresholds to significantly advance the long-standing distortion constraints from *COBE/FIRAS* (Mather et al. 1994; Fixsen et al. 1996). A particular source of SDs is due to the dissipation of ten-

---

⋆ E-mail: thomas.kite@manchester.ac.uk
† E-mail: andrea.ravenni@manchester.ac.uk
‡ E-mail: patil@lorentz.leidenuniv.nl
§ E-mail: jens.chluba@manchester.ac.uk

© 0000 RAS

sor modes while they travel almost unimpeded through the cosmic plasma (Ota et al. 2014; Chluba et al. 2015a).

How do tensor perturbations distort the CMB spectrum? In general, perturbations in the photon fluid dissipate through electron scattering and free-streaming effects. Dissipation of scalar perturbations provides one of the guaranteed sources of SDs in the early Universe within the standard thermal history (e.g., Sunyaev & Zeldovich 1970; Daly 1991; Hu et al. 1994; Chluba et al. 2012b,a). Similarly, tensor modes lose a small fraction of their energy by continuously sourcing perturbations in the photon fluid which then also distort the CMB spectrum. In contrast to scalar modes, however, the dissipation is mainly mediated by free-streaming effects. As shown in detail by Chluba et al. (2015a), this leads to dissipation of perturbations over a vast range of scales, extending far beyond those relevant to scalar perturbations. Thus, although the tensor dissipation rate is suppressed relative to scalar dissipation (tensor modes are not significantly damped by interactions with the photons), this opens new avenues for model constraints from SDs.

Building on Chluba et al. (2015a), we translate the relations between μ-distortions and primordial tensor perturbation into quantities commonly used for GW searches. This makes it easier to compare SD limits to those from other probes. As examples we consider several inflationary models which source GWs beyond vacuum fluctuations, early-universe phase transitions (PTs) and cosmic string (CS) networks, all of which demonstrate how SD measurements are and will be important for excluding portions of their respective parameter spaces. Indeed we highlight that several of the widely discussed models could have already constrained some regions of their respective parameter spaces with SD limits from *COBE/FIRAS*, taken over a quarter-century ago. Future spectrometer concepts like *PIXIE* (Kogut et al. 2016) and its enhanced versions (e.g., PRISM Collaboration et al. 2014; Kogut et al. 2019) could, through their increased sensitivity, significantly increase the range of scales and parameter space covered. This could give CMB spectral distortions an important role in this highly-synergistic multi-messenger campaign, providing unique scientific opportunities for the next generation of cosmologists and particle phenomenologists alike.

2 GWS IN THE EXPANDING UNIVERSE

A GW can be represented as a transverse traceless tensor perturbation of the metric's spatial component, h_{ij}, and the energy density it carries is $\rho_{\rm GW} = \langle h'_{ij} h'^{ij} \rangle / (32\pi G)$, where the prime denotes conformal time derivatives[1]. If these GWs were produced primordially[2], we can define the GW fractional energy density per decade of wavelengths as (e.g., Watanabe & Komatsu 2006)

$$\Omega_{\rm GW}(k,\eta) = \frac{1}{\rho_{\rm c}(\eta)} \frac{\partial \rho_{\rm GW}(k,\eta)}{\partial \ln k} = \frac{\mathcal{P}_T(k)}{12 a^2(\eta) H^2(\eta)} \left[\mathcal{T}'_{\rm GW}(k,\eta)\right]^2, \quad (1)$$

where $\rho_{\rm c}$ is the critical density, and in the second equality we factored the primordial tensor power spectrum $\mathcal{P}_T$ and the deterministic GW transfer function $\mathcal{T}_{\rm GW}$.

In Watanabe & Komatsu (2006), several analytical approximations of the GW transfer function were developed. During radiation domination (RD) we have

$$\left[\mathcal{T}'_{\rm GW}(k,\eta)\right]^2 \approx k^2 \left[j_1(k\eta)\right]^2, \quad (2)$$

whereas during matter domination (MD), one finds

$$\left[\mathcal{T}'_{\rm GW}(k,\eta)\right]^2 \approx \begin{cases} k^2 \frac{\eta_{\rm eq}^2}{\eta^2} \left[A(k) j_2(k\eta) + B(k) y_2(k\eta)\right]^2 & \text{if } k > k_{\rm eq} \\ k^2 \left[\frac{3 j_2(k\eta)}{k\eta}\right]^2 & \text{if } k < k_{\rm eq} \end{cases},$$

$$A(k) = \frac{3}{2k\eta_{\rm eq}} - \frac{\cos(2k\eta_{\rm eq})}{2k\eta_{\rm eq}} + \frac{\sin(2k\eta_{\rm eq})}{(k\eta_{\rm eq})^2}, \quad (3)$$

$$B(k) = -1 + \frac{1}{(k\eta_{\rm eq})^2} - \frac{\cos(2k\eta_{\rm eq})}{(k\eta_{\rm eq})^2} - \frac{\sin(2k\eta_{\rm eq})}{2k\eta_{\rm eq}}.$$

Here, $k_{\rm eq}$ is the comoving wavenumber entering the horizon at the time of matter-radiation equality $\eta_{\rm eq}$, and j_ℓ and y_ℓ are the spherical Bessel functions of first and second kind. For wavelengths much smaller than those entering the horizon today ($k\eta_0 \gg 1$) we can expand the GW transfer function derivatives at leading order in k. Additionally, since we always observe quantities that involve $(\mathcal{T}'_{\rm GW})^2$ integrated over some range of k, we can average over one period to obtain (e.g., Caprini & Figueroa 2018)

$$\langle [\mathcal{T}'_{\rm GW}(k,\eta)]^2 \rangle \overset{k\eta_0 \gg 1}{\underset{\approx}{\downarrow}} \eta_{\rm eq}^2 / 2\eta^4, \quad (4)$$

which is a smooth function of k valid during MD. Similarly, during RD we can apply the same procedure to Eq. (2), and obtain

$$\langle [\mathcal{T}'_{\rm GW}(k,\eta)]^2 \rangle \approx 1/2\eta^2. \quad (5)$$

For later use we point out that during RD, where Eq. (5) is valid, $a \propto \eta$, while during MD relevant to Eq. (4) we have $a \propto \eta^2$. Together with Eq. (1), this means that the GW energy density at a given scale evolves as $\Omega_{\rm GW} \propto a^{-4} H^{-2} \propto$ const during RD and $\Omega_{\rm GW} \propto a^{-4} H^{-2} \propto (1+z)$ in the MD era.

As pointed out in Watanabe & Komatsu (2006), the approximations given above neglect some important details. One of these is the process of neutrino damping, which has its greatest effects on scales important to SD physics. The damping is effective during RD but only after neutrino decoupling ($T \lesssim 2$MeV), which taken together almost exactly coincides with the SD regime. This damping occurs since free streaming neutrinos correspond to a non-negligible fraction of the energy density of the Universe during RD and generate significant anisotropic stresses that result in the damping of tensor perturbations. The magnitude of the effect is a 35.6% decrease of the power available in GW (Weinberg 2004). To include this effect the transfer function given in Dicus & Repko (2005) is used:

$$\mathcal{T}'_{\rm GW} = \frac{1}{\eta} \sum_{n \text{ even}} a_n \left[n j_n(k\eta) - k\eta j_{n+1}(k\eta)\right], \quad (6)$$

with the coefficients $a_0 = 1$, $a_2 = 0.243807$, $a_4 = 5.28424 \times 10^{-2}$ and $a_6 = 6.13545 \times 10^{-3}$. This is valid for the range of scales needed in the following section.

It is clear from Eqs. (1), (4) and (5) that the exact transfer functions are important quantities for comparing the effects of the GW background in the early and late Universe. In this paper, we compare the SD sensitivity (↔ early Universe) to PTA and interferometry (↔ late Universe). Even CMB temperature anisotropies, although sourced early on, mostly probe the Universe after the RD-MD transition. Because of this it is essential to get the exact dynamics of this transition right for any comparison to be meaningful. To study the evolution of the GW background in detail we numerically solved for the wave evolution through the RD-MD transition, which gave results agreeing with Watanabe & Komatsu (2006) and Dicus & Repko (2005) in the appropriate limits, while allowing us to more carefully model the GW background for a realistic cosmology involving neutrino and dark energy densities.

---

[1] We adopt the normalization conventions of Watanabe & Komatsu (2006).

[2] We consider the case of sub-horizon generation further on.

For the *Planck* 2018 best-fit cosmology (Aghanim et al. 2020), the exact solution is well approximated by[3]

$$\Omega_{\rm GW}/\mathcal{P}_T = \frac{\mathcal{D}\Omega_{\rm rel}}{12}\left(1+\alpha_1\kappa^{-1}+\alpha_2\kappa^{-3/2}+\alpha_3\kappa^{-2}\right), \quad \kappa = \frac{k}{k_*}, \quad (7)$$

where $\alpha_1 = 5.74$, $\alpha_2 = -2.47$, $\alpha_3 = 14.48$, $k_* = 1/551\ {\rm Mpc}^{-1}$, $\mathcal{D} = 0.642$ is the neutrino damping factor and $\Omega_{\rm rel} = 9.19\times10^{-5}$ is the combined density of radiation and neutrinos, treating the latter as massless. This solution differs by $\sim 20\%$ from Caprini & Figueroa (2018), with better matching arising when neglecting the neutrino energy density. For details see Kite et. al. (in preparation).

Since gravitational wave upper limits are usually quoted as function of frequencies rather than wavelengths, we will use the relation $k/{\rm Mpc}^{-1} = 6.5\times10^{14} f/{\rm Hz}$ to change units.

3 μ-DISTORTIONS FROM TENSOR PERTURBATIONS

Much like scalar perturbations, tensor perturbations dissipate over time, both transferring energy to neutrinos and (in smaller proportion) to photons. Primordial tensor perturbations entering the horizon during or slightly before the μ-era ($5\times10^4 \lesssim z \lesssim 2\times10^6$), when dissipating, generate μ-distortions of the CMB that will be observable today (Ota et al. 2014; Chluba et al. 2015a).

The average value of μ-distortions today is related to the primordial tensor power spectrum via a window function $W_\mu(k)$

$$\langle\mu_{\rm GW}\rangle(\eta_0) = \int {\rm d}\ln k\, W_\mu(k)\,\mathcal{P}_T(k)\,, \quad (8)$$

which already averages oscillations by integrating over a transfer function's evolution throughout the μ-era, achieving the usual factor of 1/2 implicitly. We calculate $W_\mu(k)$ numerically according to Chluba et al. (2015a). The window function is shown in Fig. 1. In comparison to the corresponding k-space window function of scalar perturbations (e.g., Chluba et al. 2012a, 2015b), the dissipation efficiency of tensors is about five orders of magnitude smaller, highlighting how weakly tensor modes couple to the photon fluid.

Offsetting this loss, we can see that tensor modes contribute to the generation of μ-distortions over a vast range of scales, with a power-law decay of contributions at $k \gtrsim 10^6\ {\rm Mpc}^{-1}$ (Fig. 1). This is in stark contrast to the dissipation of scalar perturbations, which are limited to scales $k \simeq 50 - 10{,}000\,{\rm Mpc}^{-1}$, with a strong exponential decay of contributions from $k \gtrsim 10{,}000\,{\rm Mpc}^{-1}$ (e.g., Chluba et al. 2012a). Scalar modes damp by photon diffusion, which virtually erases all perturbations once the dissipation scale is crossed. For tensors, the photon damping is minute and photon perturbations are continuously sourced by the driving tensor force, explaining this significant difference (Chluba et al. 2015a). This makes SDs a potentially unique probe of GWs from early-universe physics.

3.1 Time-dependent injection

Equation (8) determines the SD signal from primordial perturbations that were created during inflation and only later enter the horizon to dissipate their energy. Another possibility is to have perturbations created on sub-horizon scales at later times. This requires a generalisation of the window function formalism to account for the new time dependence.

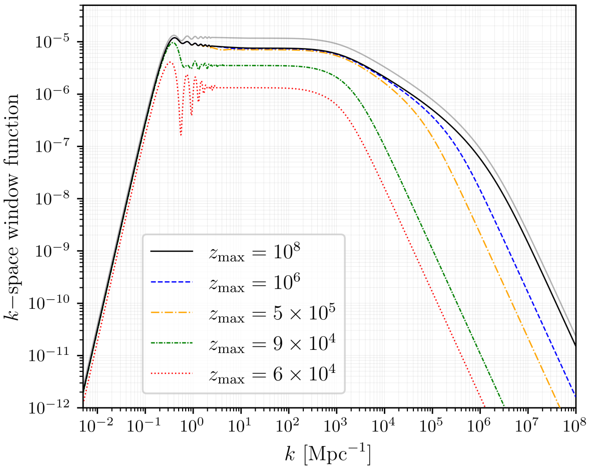


Figure 1. A series of curves demonstrating the form of the k-space window function $W_\mu^{z_{\rm max}}$ for various upper limits in redshift. For practical purposes, $z_{\rm max} = 10^8$ is equivalent to $W_\mu^\infty \equiv W_\mu$. The solid curve is the result for power injection before the μ era, while the other curves suffer from some reduced visibility at higher k. The faded line shows the results without neutrino damping, leading to a $\simeq 30\%$ increase across the window function.

An immediate difference for sub-horizon injection is that neutrino damping will not occur, as this only matters for GWs that cross the horizon between neutrino decoupling and the start of MD. This means one can use the simpler versions of the transfer function, valid in RD, given in Eqs. (2) and (5). The time dependence — which before was included in the physics underlying the window function — has to be made more explicit. Using redshift z to better match Chluba et al. (2015a), Eq. (8) can be generalized to

$$\langle\mu_{\rm GW}\rangle(z=0) = \int_0^\infty {\rm d}\ln k \int_0^\infty {\rm d}z\, \mathcal{W}_\mu(k,z)\,\mathcal{P}_T(k,z)\,, \quad (9)$$

where we introduced the GW-μ-distortion *window primitive* $\mathcal{W}_\mu(k,z)$, which captures the physics behind the damping of GWs. Note that with $\mathcal{P}_T(k,z) = \mathcal{P}_T(k)$ we recover Eq. (8) by defining $\int_0^\infty \mathcal{W}_\mu\,{\rm d}z = W_\mu$. The explicit form of the window primitive is[4].

$$\mathcal{W}_\mu = 1.4\times\frac{8H^2\eta^2}{45\dot\tau}\,[\mathcal{T}'_{\rm GW}(k,z)]^2\,\mathcal{T}_\Theta(k,z)\,{\rm e}^{-\Gamma^*_\gamma\eta}\,\mathcal{J}_\mu(z)\,, \quad (10)$$

and for convenience we summarize here the quantities which are relevant to calculate the window function (for their derivation and further explanation we refer to Chluba et al. 2015a): $\dot\tau$ is the time derivative of the Thomson optical depth. The terms $\mathcal{T}_\Theta{\rm e}^{-\Gamma^*_\gamma\eta}$ contain the physics of how the GW transfer function $\mathcal{T}_{\rm GW}$ couples to the photon fluid. These terms can be reliably approximated as

$$\mathcal{T}_\Theta(k,z) \approx \mathcal{T}_\Theta(\xi) \approx \frac{1+4.48\xi^2+91\xi^4}{1+4.64\xi+90.2\xi^2+100\xi^3+55\xi^4}\,, \quad (11a)$$

$$e^{-\Gamma^*_\gamma\eta} \approx 1\,, \quad (11b)$$

with $\xi = k/\tau'$. The final term $\mathcal{J}_\mu(\eta)$ is the energy branching ratio, which gives the fraction of total energy injected into the photon fluid that contributes to the μ distortion. We use the simple analytic approximation of the branching ratio ('method B' in Chluba 2016):

$$\mathcal{J}_\mu(z) \approx \begin{cases} {\rm e}^{-(z/z_{\rm th})^{5/2}} & \text{for } z > 5\times10^4 \\ 0 & \text{otherwise} \end{cases}\,, \quad (12)$$

with $z_{\rm th} = 1.98\times10^6$ denoting the redshift where thermalisation becomes inefficient (see also Hu & Silk 1993).

---

[3] This does **not** include a factor of 1/2 for oscillations

[4] We match the notation of Chluba et al. (2015a) with $\mathcal{T}_{\rm GW} \equiv \sqrt{2}\eta\mathcal{T}_h$

We employ one further approximation in assuming the injection happens instantaneously across all scales at a time η_*. Therefore, the tensor perturbations are uncorrelated ($\mathcal{P}_T(k, \eta < \eta_*) = 0$) up to η_* when their power spectrum abruptly jumps to some value that we will now determine. The spectrum is found at all times after η_* by redshifting $\Omega_{\rm GW}$ from the present-day value

$$\Omega_{\rm GW}(k,\eta) = \begin{cases} \frac{a^{-4}(\eta)}{E^2(\eta)}\,\Omega_{\rm GW}(k,\eta_0) & \eta > \eta_* \\ 0 & \eta < \eta_* \end{cases}, \quad E^2(\eta) \equiv \frac{H^2(\eta)}{H_0^2}, \quad (13)$$

where we used $a^{-4}/E^2 \propto \bar{\rho}_{\rm GW}/\rho_{\rm c}$.

We will only consider power injection in the RD era, hence the tensor power spectrum is then obtained using Eq. (1) together with Eq. (2), and reads

$$\mathcal{P}_T(k,\eta) = \begin{cases} \frac{12H_0^2}{a^2k^2[j_1(k\eta)]^2}\,\Omega_{\rm GW}(k,\eta_0) & \eta > \eta_* \\ 0 & \eta < \eta_* \end{cases} \quad (14)$$

Notice that if not for the fact that the tensor perturbations appear at η_*, the power spectrum would always be time-independent: it is in fact the equivalent of the primordial one in the "standard" scenario described in Chluba et al. (2015a).

Using that Eq. (14) is independent of time during RD and inserting this into Eq. (9), we can remove the time dependence in the integrand, leaving only changes in the upper limit of integration. It is therefore sufficient to study a series of window functions $W_\mu^{z_{\rm max}} = \int_0^{z_{\rm max}} \mathcal{W}_\mu \,{\rm d}z$ for different upper limits in time. Examples of $W_\mu^{z_{\rm max}}$ are shown in Fig. 1. We can observe that even modes originating from $z \gg 2 \times 10^6$ contribute to the generation of distortions. The flat plateau of the window function at $k \simeq 0.1{-}10^3\,{\rm Mpc}^{-1}$ is not affected until $z_{\rm max} \lesssim 5 \times 10^5$, and will rapidly approach $W_\mu^{z_{\rm max}} \simeq 0$ as $z_{\rm max}$ approaches 5×10^4.

For numerical applications, it is convenient to pre-tabulate the tensor window function $W_\mu^{z_{\rm max}}$ across k and injection redshift $z_{\rm max}$. Since the background cosmology is fixed to high precision (Planck Collaboration et al. 2018), this procedure avoids additional approximations.[5] However, a few comments are in place: we can further improve the treatment of the transition between the μ and y-distortion eras, which here we modelled as a step-function [see Eq. (12)]. Including the more gradual transition (e.g., see discussion in Chluba 2016), enhances the contributions from the largest scales ($k \lesssim 10^{-2}\,{\rm Mpc}^{-1}$), however, a more accurate treatment of transfer effects is also required and left to future work.

With the procedure outlined in this section we can calculate the tensor dissipation contribution to the present day value of μ-distortions. However, other processes, such as dissipation of acoustic modes and Compton cooling also source μ-distortions, henceforth referred to as $\mu_{\rm other}$. Any non-detection of an enhanced level of SD would straightforwardly constrain models that generate a large $\mu_{\rm GW}$, comparable to or greater than $\mu_{\rm other}$. However, things are more delicate when $\mu_{\rm GW}$ becomes much smaller than the value of $\mu_{\rm other}$ expected in the standard cosmological model, $\mu_{\rm other} \simeq 2 \times 10^{-8}$ (e.g., Chluba 2016). In this regime, any actual analysis would require a marginalization of other sources, that we do not take into account here. However, assuming standard slow-roll inflation, we can in principle accurately predict the expected standard contribution given the power spectrum parameters measured at large angular scales (Chluba et al. 2012b; Khatri & Sunyaev 2013; Chluba & Jeong 2014; Cabass et al. 2016; Chluba 2016). For simplicity, we shall thus assume perfect removal of other μ-contributions.

Below we will consider the upper limit on μ-distortions set by *COBE/FIRAS* ($\mu < 9 \times 10^{-5}$ 95%CL) (Mather et al. 1994; Fixsen et al. 1996), and the forecasted constraints for *PIXIE* ($\mu < 3 \times 10^{-8}$) (Kogut et al. 2011), *SuperPIXIE* ($\mu < 7.7 \times 10^{-9}$) (Kogut et al. 2019), *Voyage 2050* ($\mu < 1.9 \times 10^{-9}$) and 10×*Voyage 2050* ($\mu < 1.9 \times 10^{-10}$) (Chluba et al. 2019b), all of which already account for the presence of foregrounds following Abitbol et al. (2017).

3.2 Scalar contributions

Above we discussed separating $\mu_{\rm GW}$ from $\mu_{\rm other}$, taking the latter to be the standard model expected value. A second discussion is necessary, however, regarding the contribution that scalar perturbations have to a μ signal. This is important, considering that energetic early-universe phenomena have the potential to generate scalar perturbations as well as tensors, which will enhance the SD production. In the following section we will discuss the scalar contributions for models where it is possible to do so, but some statements apply in general: Chluba et al. (2015a) show that the corresponding window functions for scalar perturbations peak around 10^5 higher than for tensors, but for a narrower range of scales ($k \simeq 50 - 10{,}000\,{\rm Mpc}^{-1}$ or $f \simeq 8 \times 10^{-14} - 1.5 \times 10^{-11}\,{\rm Hz}$, as previously discussed). Thus, for tensor perturbations to dominate the spectral distortion signal the scalar spectrum created must be less than 1 part in 10^5 of the tensor spectrum, or must be injected on smaller scales than $k \sim 10^4{\rm Mpc}^{-1}$. Provided both the wider tensor window, and that some early processes will be almost invisible to scalar probes, the machinery explained above for constraining early tensor energy injection are still of interest and importance, despite the relatively low sensitivity.

4 MINIMALLY PARAMETRIC CONSTRAINTS

In this section, we calculate the constraining power of spectroscopic CMB measurements in a minimally parametric fashion. As in Campeti et al. (2021), we parametrize the primordial tensor power spectrum using logarithmically spaced tophat functions centered around some $\ln k_i$ with $\ln k_{i+1} - \ln k_i = 1.2\,\forall i$. This allow us an easy comparison with Fig. 8 of their paper:

$$\mathcal{P}_T(k) = \sum_i A_i W_i(k)\,, \quad (15a)$$

$$W_i(k) = \begin{cases} 1 & \text{if } \ln k \in [\ln k_i - 0.6, \ln k_i + 0.6] \\ 0 & \text{otherwise} \end{cases}. \quad (15b)$$

Therefore, for each i, we insert Eq. (15a) into Eq. (8), and calculate the maximum value of A_i that is compatible with the chosen $\langle \mu_{\rm GW} \rangle(\eta_0)$ upper limit. With that information we then use Eq. (1) to calculate the corresponding $\Omega_{\rm GW}$ constraint.

In Fig. 2 we show the sensitivity curves for *COBE/FIRAS*, *PIXIE*, *SuperPIXIE*, *Voyage 2050* and 10×*Voyage 2050*, which all include estimated penalties from foregrounds. For comparison, we also report the sensitivity curves from Campeti et al. (2021), which recently compiled the results of many planned experiments (Hazumi et al. 2019; Smith & Caldwell 2019; Arca Sedda et al. 2021; Sesana et al. 2019; Kuroyanagi et al. 2015; Crowder & Cornish 2005; El-Neaj et al. 2020; Reitze et al. 2019; Hild et al. 2011; Weltman et al. 2020). Moreover, we show the NANOGrav 12.5 year observation (Arzoumanian et al. 2020), interpreted as GW stochastic background according to their 5 frequency power-law model (see Kuroyanagi et al. 2020, for more discussion on whether the signal can be inflationary). Since the extrapolation of a red spectra would be favourable for a SD detection, we conservatively assume a flat spectrum.

---

[5] A simple interpolation routine to calculate $W_\mu^{z_{\rm max}}$ is available here: `https://github.com/CMBSPEC/GW2SD.git`

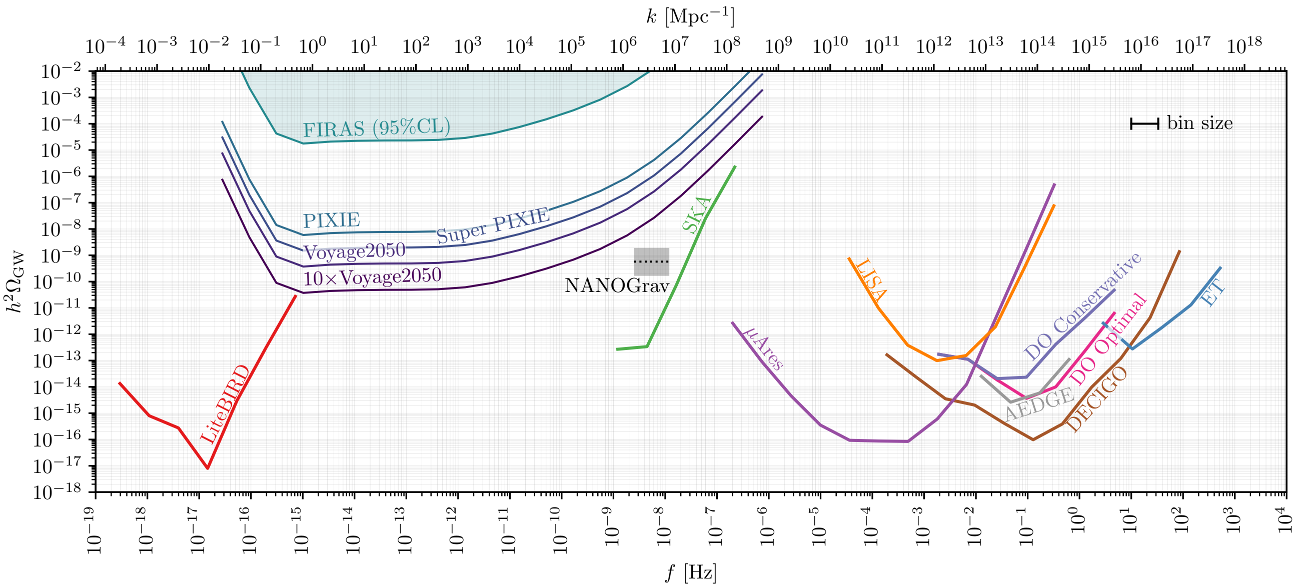


Figure 2. Upper limits on the energy density of gravitational waves from measurements of μ-distortions for various experimental configurations (*COBE/FIRAS*, *PIXIE*, *SuperPIXIE*, *Voyage 2050*, 10×*Voyage 2050*). For ease of comparison, we also report the upper limits for various other CMB, PTA and direct detection experiments (taken from Campeti et al. 2021), and the NANOGrav 12.5 years 95% confidence interval assuming a flat spectrum.

While the existing constraint derived from the *COBE/FIRAS* data is a few order of magnitude higher than other probes, next generation satellites will start to bridge nicely the frequency gap existing between CMB observation and direct GW detection. It is interesting to notice that the upper bound from SDs will cover a very broad range of frequencies (more than 5 decades in f). As such, any signal that is not sharply peaked in frequency will generate a comparatively higher μ-distortion, tightening the constraints on specific parametric models, as we will see in the next section.

5 CONSTRAINTS ON SPECIFIC MODELS

In this section, we consider concrete models that generate GWs over a wide range of scales. For each of the following models it is enough to insert their corresponding tensor power spectrum into Eq. (8) or (9) to obtain the predicted μ-distortion, depending on whether the injection is primordial or happens after reheating.

Generally speaking, once accounting for the limits on r from *Planck* (Ade et al. 2018; Aghanim et al. 2020), we understand that appreciable SDs can only be created by models with substantially enhanced tensor power at small scales. To also avoid future constraints at small scales, models with localized features at $f = 10^{-15} - 10^{-9}$ Hz are most promising. In the context of PTs, for example, this identifies low-scale dark or hidden sector transitions at energies $\simeq$ 10 MeV - 10 eV in the post-inflation era as a target.

5.1 Single-field slow-roll inflation

As a benchmark we consider the tensor perturbations generated by single-field slow-roll inflation. This model predicts a very low, almost scale invariant tensor spectrum, and as such we cannot expect SD constraints to be competitive with either CMB measurements or future direct detections at small scales. We however include the model for completeness, and as a point of comparison. The tensor spectrum from this model is given by

$$\mathcal{P}_T^{\rm sf} = A_T\,(k/k_0)^{n_T}\,, \tag{16}$$

where the amplitude of tensor and scalar perturbations A_T and A_S are related by the tensor to scalar ratio by $r \equiv A_T/A_S$, and $n_T = -r/8$ (Lyth & Riotto 1999). Current constraints, mostly driven by *Planck* low-ℓ temperature and BICEP2/Keck B-modes data, (Ade et al. 2018; Aghanim et al. 2020) set the upper limit $r_{0.002} < 0.06$ (95%) at $k = 0.002\,{\rm Mpc}^{-1}$. Upon noticing $|n_T| \leq 0.0075 \approx 0$, one can approximate the result by integrating a flat spectrum yielding $\langle\mu\rangle(r) \approx 1.68 \times 10^{-13}\, r$ which gives the correct result to within $\leq 5\%$ for all values not ruled out by *Planck*. This shows that for any allowed value of r the SD signal will be out of reach for even the most sensitive SD mission concepts.

In principle this contribution is present as a component of tensor spectrum in the other models considered in the following sections. However, since the amount of SD it generates is anyway negligible, we will omit it in the following. Note that the *Planck* constraint on r will also be considered for other models. Strictly speaking, the aforementioned constraint only apply to a power-law tensor spectrum, a condition not necessarily met by the models we will consider in the following. To provide some context to the SD constraints we will draw, we opt to employ an order-of-magnitude estimate of the *Planck* constraint, simply requiring that any spectrum of tensor perturbations, $\mathcal{P}_T(k)$, must satisfy $\mathcal{P}_T(k)/\mathcal{P}_S^{\rm sf}(k)\big|_{k=0.002\,{\rm Mpc}^{-1}} < 0.06$. In principle, a proper analysis of the *Planck* and BICEP2/Keck data could be carried out to set constraints on the models that will be discussed here. This, however, goes beyond the scope of the paper. *LiteBIRD* (Hazumi et al. 2019), providing low multipole *BB* information at much higher precision, will allow us to further improve the limits set by *Planck* on the same range of scales in the near future.

5.2 Spectator SU(2) axions

Many inflationary models require the dynamics of additional spectator fields active during the inflationary period, itself driven by a separate scalar field. Generally speaking, the dynamics of the spectator field generate tensor perturbations in addition to those produced by the vacuum fluctuations of the quasi de Sitter background.

In this section, following Campeti et al. (2021), we consider an axion-SU(2) spectator field based on the "chromo-natural" inflation model (Adshead & Wyman 2012; Dimastrogiovanni et al. 2017). Here, the SU(2) gauge fields acquire an expectation value, the fluctuations around which include a tensor perturbation with a bilinear coupling to the graviton. The dynamics of the spectator SU(2) axion are such that gravitons of a particular helicity are amplified via a transient tachyonic instability, resulting in a (circularly polarized) contribution to the tensor power spectrum.

The spectrum for this model is given by (see Thorne et al. 2018)

$$\mathcal{P}_T^{\rm SU(2)}(k) = r_* \mathcal{P}_{\mathcal{R}}^{\rm sf}(k) \exp\left[-\frac{1}{2\sigma^2}\ln^2\left(\frac{k}{k_p}\right)\right], \tag{17}$$

which relates to the spectrum of scalar perturbations

$$\mathcal{P}_{\mathcal{R}}^{\rm sf} = A_S\,(k/k_0)^{ns-1}\,. \tag{18}$$

In order to constrain this model, we use the best-fit *Planck* parameters (Planck Collaboration et al. 2018) for Eq. (18). However, r_*, k_p and σ are related to the parameters of the gauge theory and are essentially free to vary here. We take as a reasonable set of values, those given in Campeti et al. (2021) (their model AX3) $(r_*, k_p, \sigma) = (50, 10^6\ {\rm Mpc}^{-1}, 4.8)$, which would yield $\mu = 2.1\times10^{-12}$. Entertaining the question as to which set of parameters would maximize the μ distortion signal while satisfying both observational and model constraints we find $(r_*, k_p, \sigma) = (265, 2.85\times10^4\ {\rm Mpc}^{-1}, 4.02)$. However, even this best case scenario leaves no appreciable SD signal, yielding $\mu = 2.1\times10^{-11}$. This result can be understood by considering the parabolic shape of the spectrum in log$\mathcal{P}$-log$k$ space, which due to model constraints cannot peak to sharply (e.g. see Eqs. (A8) and (A11) in Thorne et al. 2018). This means that a spectrum which avoids the *Planck* constraints cannot simultaneously peak too high in the SD regime. In contrast, the models considered in the following subsections have spectra resembling broken power laws, and can be much more effective in satisfying current constraints while simultaneously generating significant SDs.

5.3 Ultra-light U(1) audible axions

The sensitivity of SD measurements to axion standard model extensions have already been discussed in the literature (Mukherjee et al. 2018). In this subsection however we consider a model proposed in Machado et al. (2019); Machado et al. (2020), in which the axions specifically produce a strong GW signal. In this scenario (generic in the context of string compactifications), one has the presence of one or more U(1) axions with mass m and decay constant f_ϕ that couple to dark sector photons. At early times during radiation domination, when the Hubble parameter H is greater than m, the axion field is over-damped and is frozen. Once $H \lesssim m$, corresponding to the temperature $T \approx \sqrt{mM_{\rm pl}}$, the axion starts to oscillate around the minimum of its potential, sourcing gauge field production of a particular helicity that goes on to generate GWs. Since these GWs are only produced on sub-horizon scales after the axion starts oscillating, the results of Sect. 3.1 are essential in finding the μ signal accurately.

The audible axion scenario features a qualitative difference with models that secondarily generate gravitation waves via gauge field production during inflation, as the generation occurs during radiation domination, when $H < m$. The oscillating axion at the minimum of the potential must remain a sub-dominant contribution to the energy budget, otherwise we'd have a phase of intermediate matter domination.

It follows that the contribution of the oscillating axion and the subsequently produced dark photons must be sufficiently subleading to the energy density in the radiation fluid. Their relative contributions to the curvature perturbation in total comoving gauge will consequently be suppressed relative to the contribution from fluctuations in the radiation fluid already present before dark photon production (originating from the vacuum fluctuations sourced during inflation). Hence the contribution of scalar perturbations sourced by axion dynamics to the $\mu-$distortion signal will be subleading to those generated by primordial perturbations from inflation, and can safely be neglected.

This model is of particular interest to us as it produces a narrower spectrum of GWs. Thus, to constrain its parameter space it is important to have probes that can cover all phenomenologically relevant frequencies. The GWs produced can be parametrized as a spectrum of the form

$$\Omega_{\rm GW}^{\rm U(1)}(k) = \frac{6.3\,\Omega_{\rm GW}^{\rm U(1)}(f_{\rm AA})\left(k/\tilde{k}\right)^{1.5}}{1+\left(k/\tilde{k}\right)^{1.5}\exp\left[12.9\left(k/\tilde{k}-1\right)\right]}\,, \tag{19a}$$

with

$$\tilde{k} = 1.3\times10^{15}\ [f_{\rm AA}/{\rm Hz}]\ {\rm Mpc}^{-1}. \tag{19b}$$

Here $\Omega_{\rm GW}(f_{\rm AA})$ and $f_{\rm AA}$ are a function of the free parameters of the model. These parameters, as introduced in Machado et al. (2019); Machado et al. (2020), are f_ϕ, m, α and θ, relating to the *fit parameters* in Eq. (19a) via

$$f_{\rm AA} \approx 6\times10^{-4}\ {\rm Hz}\left[\frac{\alpha\theta}{66}\right]^{2/3}\left[\frac{m}{10{\rm meV}}\right]^{1/2}, \tag{20a}$$

$$\Omega_{\rm GW}^{\rm U(1)}(f_{\rm AA}) \approx 1.67\times10^{-4} g_{\rho,*}^{-1/3}\left[\frac{f_\phi}{M_{\rm pl}}\right]^4\left[\frac{\theta^2}{\alpha}\right]^{4/3}, \tag{20b}$$

which have both been redshifted to their present-day values. The first two free parameters (i.e., f_ϕ and m) essentially dictate the height and frequency of the peak in the power spectrum respectively. The second two parameters are limited to $\alpha \sim 10-100$ and $\theta \sim \mathcal{O}(1)$, and do not significantly change the shape of the spectrum for the range of allowed values. These parameters are therefore degenerate with the first two. We choose fiducial values of $\alpha = 60$ and $\theta = 1$, but the main results given here hold more generally.

The direct dependence of $f_{\rm peak}$ on m means that different types of experiment will probe different mass scales. This is shown Fig. 3, where vertical dotted lines distinguish where different detection methods are dominant. From here it can be seen that SD are sensitive to the ultralight limit of the $U(1)$ audible axion model, a result which again holds for any valid combination of α and θ.

Note that *Planck* extends the limits from *COBE/FIRAS* at low masses to smaller values of f_ϕ. Future SDs measurements could significantly improve the limits from *Planck* to higher masses, covering a wider range of the parameter space of phenomenological interest.

We note in particular how SDs can constrain masses in a range not accessible to other measurements ($10^{-22}-10^{-13}$ eV). Such ultralight axions may be ubiquitous in particular string compactifications (Arvanitaki et al. 2010), and moreover, could be a viable dark matter candidate were they to form a condensate at late times (Hui et al. 2017; Marsh 2016), further illustrating the utility of SDs for particle phenomenology.

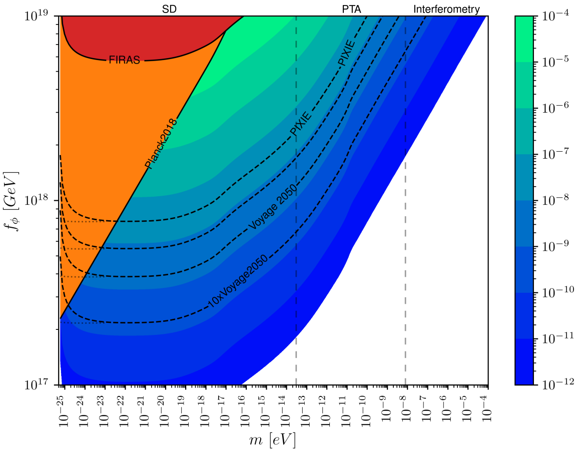


Figure 3. A contour plot showing the expected SD signal arising from different combinations of f_ϕ and m in the U(1) model. Without loss of generality, fiducial values of $\alpha = 60$ and $\theta = 1$ were chosen. Contours showing the visibility of several proposed spectrometers are shown. Vertical dotted lines indicate regions of the phase space where different probes are most sensitive (from left to right: SD, PTA and Interferometry). Dotted lines continuing the SD mission contours show the estimates ignoring late time injection.

5.4 Phase transitions beyond the Standard Model

The post-inflationary epoch may have seen a variety of first order phase transitions (PTs) in theories that go beyond the standard model of particle physics. First order PTs are characterized by the fact that latent energy is released, and phases of true vacuum nucleate within false vacuum domains, resulting in bubble collisions (BC) that generate a stochastic GW background. Moreover, magneto-hydrodynamic (MHD) turbulence and sound waves (SW) in the bulk plasma during and after the phase transition also source sub-horizon GWs at commensurate frequencies. If these processes take place during the μ-era or shortly before, they can potentially result in measurable SDs. Here we once again use the results of Sect. 3.1 to calculate the associated SDs.

Referring to the review of Caprini & Figueroa (2018), we see that the spectra resulting from the three different mechanisms for GW production from PTs are given by

$$h^2\Omega_{\rm GW}^{\rm BC}(f) = 1.67\times10^{-5}\left(\frac{H_*}{\beta}\right)^2\left(\frac{\kappa_{\rm BC}\alpha}{1+\alpha}\right)^2\left(\frac{100}{g_*(T_*)}\right)^{\frac{1}{3}} \qquad (21a)$$
$$\times\left(\frac{0.11v_w^3}{0.42+v_w^2}\right)\frac{3.8(f/f_{\rm BC})^{2.8}}{1+2.8(f/f_{\rm BC})^{3.8}}\,,$$

$$h^2\Omega_{\rm GW}^{\rm SW}(f) = 2.65\times10^{-6}\left(\frac{H_*}{\beta}\right)\left(\frac{\kappa_v\alpha}{1+\alpha}\right)\left(\frac{100}{g_*(T_*)}\right)^{\frac{1}{3}} \qquad (21b)$$
$$\times\, v_w\left(\frac{f}{f_{\rm SW}}\right)^3\left(\frac{7}{4+3(f/f_{\rm SW})^2}\right)^{\frac{7}{2}}\,,$$

$$h^2\Omega_{\rm GW}^{\rm MHD}(f) = 3.35\times10^{-4}\left(\frac{H_*}{\beta}\right)\left(\frac{\kappa_{\rm MHD}\alpha}{1+\alpha}\right)^{\frac{3}{2}}\left(\frac{100}{g_*(T_*)}\right)^{\frac{1}{3}} \qquad (21c)$$
$$\times\, v_w\frac{(f/f_{\rm MHD})^3}{[1+(f/f_{\rm MHD})]^{\frac{11}{3}}\,(1+8\pi f/h_*)}\,,$$

with peak frequencies

$$\chi_0 = \left[\frac{\beta}{H_*}\right]\left[\frac{T_*}{100{\rm GeV}}\right]\left[\frac{g_*(T_*)}{100}\right]^{\frac{1}{6}}\,, \qquad (21d)$$

$$f_{\rm BC} = 1.65\times10^{-5}\,{\rm Hz}\left(\frac{0.62}{1.8-0.1v_w+v_w^2}\right)\chi_0\,, \qquad (21e)$$

$$f_{\rm SW} = 1.9\times10^{-5}\,{\rm Hz}\,v_w^{-1}\,\chi_0\,, \qquad (21f)$$

$$f_{\rm MHD} = 2.7\times10^{-5}\,{\rm Hz}\,v_w^{-1}\,\chi_0\,. \qquad (21g)$$

Here, the three principal model parameters are α, β and v_w, which fix the amount of latent heat released by the transition as a fraction of the total energy density, inverse time duration of the PT, and velocity of bubble walls respectively. Denoted with $*$ are quantities at the time of the PT, making another key parameter $z_{\rm PT}$. The first two parameters follow $0 \leq \alpha \leq 1$ and $\beta/H_* > 1$.

The velocity of sound waves has been set to unity, since bubble walls usually propagate close to the speed of light. Parameters labelled $\kappa_i \in [0,1]$ give the weighted contribution from each mechanism. For this work we have used $\kappa_{\rm BC} = 1$, $\kappa_{\rm MHD} = \kappa_v$ and

$$\kappa_v \approx \frac{\alpha}{0.73+0.083\sqrt{\alpha}+\alpha}, \qquad (22)$$

the last of which is valid for $v_w \simeq 1$. The expected SD limits on PTs given these considerations are shown in Fig. 4. Even for low-energy PTs ($\alpha = 0.1$) a *PIXIE*-like mission would explore some of the parameter space not already excluded by *Planck*; however, it would only see rather long PT. In the more energetic cases ($\alpha \geq 0.5$), SD missions could realistically detect PT lasting small fractions of the age of the Universe, and occurring relatively late in cosmic history.

Evidently, SDs provide a unique and complimentary window into low scale phase transitions (corresponding to energy scales in the range 10 Mev - 10 eV) that are not possible to probe with any other observation. An important caveat to our discussion of this scenario is the potential for the generation of sub-horizon scalar perturbations during and after phase transitions. Sub-dominant contributions arising from the scalar field dynamics have been calculated in Cutting et al. (2018), however retaining the scalar contributions from sound waves and MHD turbulence generated after the transition will require further study, and remains an important open question for the present analysis. Our limits can therefore be considered conservative.

5.5 GUT cosmic string networks

Another tell tale sign of physics beyond the Standard Model is the existence of topological defects. Excluding textures, the standard model does not allow for any defects. However, larger gauge symmetries (ubiquitous in models that go beyond the Standard Model) could admit symmetry breaking patterns that generate topological defects in the early Universe (see Kibble 1980, 1982) which could have persisted into cosmologically observable epochs. Although the simplest models of monopoles and domain walls are tightly constrained (see Sects. 13.5.3 and 14.3.3 in Vilenkin & Shellard 1994), cosmic strings networks remain a theoretical possibility and can impart potentially observable GW signals (see Sect. 10.4 of Vilenkin & Shellard 1994).

As an example, we consider a model proposed by Buchmuller et al. (2019) which attempts leptogenesis within an SO(10) grand unified theory via a $\mathrm{U}(1)_{B-L}$ phase transition, where a local U(1) baryon minus lepton number symmetry is spontaneously broken.

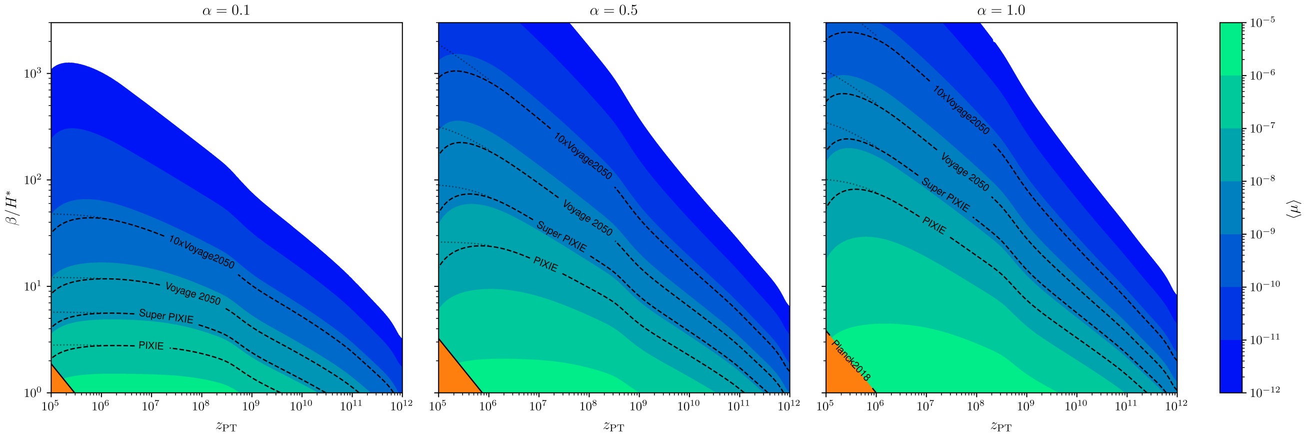


Figure 4. A series of contour plots showing the expected SD signal arising from low scale first order phase transitions. Dotted lines (visible on the left) give the sensitivity with standard window function, $W_\mu(k)$, showing that late power injection leads to a decrease of less than an order of magnitude for $10^5 < z_{\rm PT} < 10^6$. The limits from *Planck* are also shown for comparison. The temperature at the time of the PT can be found with $T_{\rm PT}/{\rm MeV} \approx 2\times 10^{-10} z_{\rm PT}$.

The result of the $B-L$ transition will be a meta-stable CS network generated at the time of the transition, which over the course of the collapse generates a mostly flat spectrum of GWs due to the decay of string loops. An approximate form of their spectrum is given in terms of the model parameters κ and $G\mu$ as[6]

$$h^2\Omega_{\rm GW}^{\rm CS} = h^2\Omega_{\rm GW}^{\rm plateau}\min\left[(f/f_*)^{3/2}, 1\right], \tag{23a}$$

$$f_* = 3\times10^{14}\,{\rm Hz}\,{\rm e}^{-\pi\kappa/4}\left[\frac{G\mu}{10^{-7}}\right]^{-1/2}, \tag{23b}$$

$$h^2\Omega_{\rm GW}^{\rm plateau} = 8.04\,\Omega_{\rm r}h^2\left[\frac{G\mu}{\Gamma}\right]^{1/2}. \tag{23c}$$

Buchmuller et al. (2019) give a value of $\Gamma\approx 50$ for this particular model, and we use a value of $\Omega_{\rm r}h^2 = 2.5\times10^{-5}$.

In reality string network collapse would be a function of time, but to match the formalism outlined in Sect. 3.1 we conservatively assume the entire spectrum emerges at the final moment of collapse given by Buchmuller et al. (2019)

$$z_{\rm collapse} = \left(\frac{70}{H_0}\right)^{1/2}\left(\Gamma\,\frac{(G\mu)^2}{2\pi G}\,e^{-\pi\kappa}\right)^{1/4}. \tag{24}$$

The spectrum grows $\propto f^{3/2}$ up to f_*, and is flat for higher frequencies. Furthermore, f_* only depends weakly on $G\mu$ but varies significantly with κ. This means that once κ is large enough that the spectrum is flat across the entire window of visibility for a given experiment, the probe will only be sensitive to $G\mu$. With SD missions probing lower frequencies than astrophysical probes they will be complementary in limiting the lower bounds of the κ parameter. The potential of SD missions for constraining this model is shown in Fig. 5.

Given that the GW spectra produced by CS network collapse has a plateau at smaller scales, for any given sensitivity depicted in Fig. 2, we see that any one of the probes depicted will be equally good at detecting the GW background produced. It is also worth noting that the type of spectrum considered here will hold more generally for a wide range of CS models (see Figueroa et al. 2020).

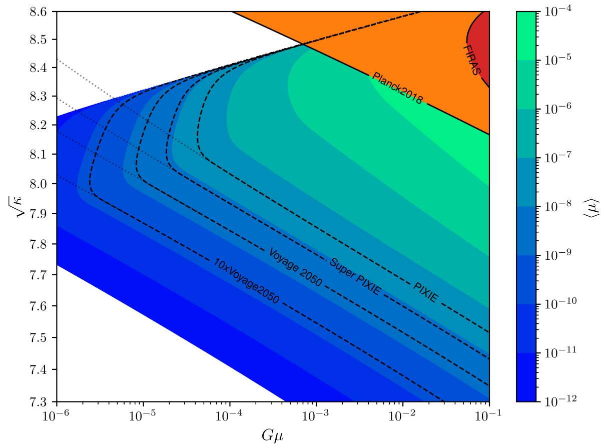


Figure 5. A contour plot showing the expected μ signal from a CS network arising from a $\mathrm{U}(1)_{B-L}$ phase transition at the GUT scale. The limit placed by *COBE/FIRAS* is shown in red, and similarly for *Planck* in orange. Dashed contours show the sensitivity of various proposed SD missions. Faint dotted lines show the contours without using the $z_{\rm max}$ limited window function.

Although the generation of scalar perturbations of CS networks in the scaling regime is well understood, the situation is much less clear for the scalar perturbations generated from the decay of meta-stable networks. Moreover, it is unclear whether the dominant decay channel will be via gravitational wave production or scalar perturbations, and the answer will certainly be very model dependent. This remains another open question as far as this study is concerned, and the absence of any such detailed calculations is itself perhaps due to the fact that it may have been unclear in the past what observational consequences, if any, sub-horizon generation of scalar perturbations generated by CS network collapse would have. We hope the results of this paper will provide the necessary motivation for such calculations.

---

[6] Not to be confused with the SD amplitude μ. The combination $G\mu$ will always be in reference to the energy scale of the CS physics.

6 DISCUSSION AND CONCLUSIONS

Highly energetic events in the early Universe, either during inflation or subsequently during radiation domination, can inject power into the GW spectrum. This can include GWs from sources within the standard ΛCDM cosmology, or from models that invoke physics beyond it. Detecting the GW spectrum is therefore key to further scrutinizing our current paradigm, as well as pushing our knowledge of the early Universe to new and exciting areas. Future experiments will probe these stochastic backgrounds, each sensitive to a range of frequencies/wavelengths dictated by the nature of the experiment. As we have highlighted here, a wide range of GW frequencies ($f = 10^{-15}$–10^{-9} Hz) can only be probed by SD observations. This large span of wavelengths compensates for the relatively low efficiency of generating SDs from GWs, thus making them a potentially powerful probe of physics beyond the Standard Models of both particle physics and cosmology.

This work aims to introduce SDs as a complimentary probe through which one can detect and constrain stochastic GW backgrounds. The fundamental element to link these two messengers is the k-space window function, which maps a given GW spectrum into a SD signal imprinted before last scattering [see Eqs. 8 and 9]. In order to study the injection of power on sub-horizon scales, the window function for primordial tensor perturbations has been generalised [see Eq. 9], leading to minor changes in some models (Fig. 3) but large changes in others (Fig. 5). This is essentially related to the fact that GWs have less cosmic history to dissipate their energy to the photon-baryon plasma. A simple python tool is provided at `GW2SD`[7] and allows one to easily estimate SD limits on various models, given the tensor power spectrum, $\mathcal{P}_T(k, z)$, that comes into existence at a single redshift z. This is certainly a good approximation for 1'st order phase transitions, and holds to a good approximation for scenarios that dynamically generate GWs over a short duration. Refinements to account for the exact time-dependence of the process are left to future work.

To illustrate the utility of SDs for GW cosmology, a series of phenomenological models were discussed, and their resulting SD signals studied: As expected, the tensor perturbations generated by single-field slow-roll inflation are too weak to be measured with SDs (Ota et al. 2014; Chluba et al. 2015a). Spectator axion-SU(2) fields too, even in more favourable cases that we considered, will realistically be out of reach in the foreseeable future. The Audible axion model (Sect. 5.3) on the other hand, can have a large region of its parameter space constrained by SDs, particularly for a wide range of masses in the ultra-light regime (Fig. 3). Similarly, the GWs from low scale (10 eV - 10 MeV) dark sector phase transitions in the early Universe will be visible with future SD missions if the relative energy content of the participating field is sufficiently large, and the duration sufficiently long (see Sect. 5.4 and Fig. 4). The typically flat GW spectra produced by CS networks can be seen by many instruments, but SDs will be complementary to other probes in being sensitive to string collapse especially in the μ-era. It is noteworthy that some of the aforementioned models were already constrained with *COBE/FIRAS* long before first limits from *Planck* existed. Future CMB spectrometers like *SuperPIXIE* (Kogut et al. 2019) could establish a new frontier in this respect.

As we have discussed, both in general and for specific models, any SD constraints should include both the scalar and tensor perturbations arising from energetic events (e.g. see Tashiro et al. 2013; Amin & Grin 2014, for SDs from scalar perturbations of CS and PTs, respectively). This potential for combining sources is another advantage SD experiments have over GW-based experiments, since the latter are only sensitive to the direct tensor perturbations. This advantage is not easily utilised for the models discussed in this paper, however, since the necessary scalar spectra are generally absent from the literature. Where the inclusion *is* possible, it is again important to highlight that SDs from tensor perturbations cover a wider range of physical scales than SDs from scalar sources, thus extending the reach of SDs to earlier epochs. In addition, some scenarios do not produce any significant scalar perturbations [e.g., the axion-SU(2) model], making it crucial to account for SDs caused by tensor perturbations. Overall, SDs uniquely probe the presence of small-scale perturbations in regimes that are not directly accessible, thus highlighting the important role that future CMB spectrometers could play in GW cosmology, and, by extension, beyond the Standard Model phenomenology.

DATA AVAILABILITY

Window functions (e.g. Fig. 1) are available at `https://github.com/CMBSPEC/GW2SD.git`.

ACKNOWLEDGMENTS

We would like to thank Ana Achucarro and Jose Juan Blanco-Pillado for discussions about scalar perturbations generated during cosmic string collapse. We also thank Paolo Campeti and Eiichiro Komatsu for providing the data used in Fig. 2, together with helpful discussions regarding the mapping between $\mathcal{P}_T$ and $\Omega_{\rm GW}$. This work was supported by the ERC Consolidator Grant *CMBSPEC* (No. 725456) as part of the European Union's Horizon 2020 research and innovation program. TK was further supported by STFC grant ST/T506291/1. JC was also supported by the Royal Society as a Royal Society URF at the University of Manchester, UK.

REFERENCES

Abbott B., et al., 2019, Phys. Rev. X, 9, 031040
Abbott B. P., et al., 2020a, The Astrophysical Journal Letters, 892, L3
Abbott R., et al., 2020b, The Astrophysical Journal Letters, 896, L44
Abitbol M. H., Chluba J., Hill J. C., Johnson B. R., 2017, Monthly Notices of the Royal Astronomical Society, 471, 1126–1140
Ade P., et al., 2018, Phys. Rev. Lett., 121, 221301
Ade P., et al., 2019, JCAP, 02, 056
Adshead P., Wyman M., 2012, Phys. Rev. Lett., 108, 261302
Aghanim N., et al., 2020, Astron. Astrophys., 641, A6
Alam M. F. et al., 2020, arXiv e-prints, arXiv:2005.06490
Amin M. A., Grin D., 2014, Physical Review D, 90, 083529
Arca Sedda M. et al., 2021, arXiv e-prints, arXiv:2104.14583
Arvanitaki A., Dimopoulos S., Dubovsky S., Kaloper N., March-Russell J., 2010, Phys. Rev. D, 81, 123530
Arzoumanian Z. et al., 2020, The Astrophysical Journal Letters, 905, L34
Buchmuller W., Domcke V., Murayama H., Schmitz K., 2019, arXiv e-prints, arXiv:1912.03695
Burigana C., Danese L., de Zotti G., 1991, Astronomy & Astrophysics, 246, 49
Cabass G., Melchiorri A., Pajer E., 2016, Physical Review D, 93, 083515
Campeti P., Komatsu E., Poletti D., Baccigalupi C., 2021, Journal of Cosmology and Astroparticle Physics, 2021, 012

---

[7] `https://github.com/CMBSPEC/GW2SD.git`

Caprini C., Figueroa D. G., 2018, Class. Quant. Grav., 35, 163001

Chluba J., 2016, Monthly Notices of the Royal Astronomical Society, 460, 227

Chluba J. et al., 2019a, arXiv e-prints, arXiv:1909.01593

Chluba J., Dai L., Grin D., Amin M. A., Kamionkowski M., 2015a, Monthly Notices of the Royal Astronomical Society, 446, 2871

Chluba J., Erickcek A. L., Ben-Dayan I., 2012a, The Astrophysical Journal, 758, 76

Chluba J., Hamann J., Patil S. P., 2015b, International Journal of Modern Physics D, 24, 1530023

Chluba J., Jeong D., 2014, Monthly Notices of the Royal Astronomical Society, 438, 2065

Chluba J., Khatri R., Sunyaev R. A., 2012b, Monthly Notices of the Royal Astronomical Society, 425, 1129

Chluba J. et al., 2019b, BAAS, 51, 184

Chluba J., Sunyaev R. A., 2012, Monthly Notices of the Royal Astronomical Society, 419, 1294

Crowder J., Cornish N. J., 2005, Phys. Rev. D, 72, 083005

Cutting D., Hindmarsh M., Weir D. J., 2018, Phys. Rev. D, 97, 123513

Daly R. A., 1991, The Astrophysical Journal, 371, 14

Danese L., de Zotti G., 1982, Astronomy & Astrophysics, 107, 39

Delabrouille J. et al., 2019, arXiv e-prints, arXiv:1909.01591

Dicus D. A., Repko W. W., 2005, Physical Review D, 72, 088302

Dimastrogiovanni E., Fasiello M., Fujita T., 2017, JCAP, 01, 019

El-Neaj Y. A., et al., 2020, EPJ Quant. Technol., 7, 6

Figueroa D. G., Hindmarsh M., Lizarraga J., Urrestilla J., 2020, Physical Review D, 102, 103516

Fixsen D. J., Cheng E. S., Gales J. M., Mather J. C., Shafer R. A., Wright E. L., 1996, The Astrophysical Journal, 473, 576

Hazumi M., et al., 2019, J. Low Temp. Phys., 194, 443

Hild S., et al., 2011, Class. Quant. Grav., 28, 094013

Hu W., Scott D., Silk J., 1994, The Astrophysical Journal Letters, 430, L5

Hu W., Silk J., 1993, Physical Review D, 48, 485

Hui L., Ostriker J. P., Tremaine S., Witten E., 2017, Phys. Rev. D, 95, 043541

Illarionov A. F., Sunyaev R. A., 1975, Soviet Astronomy, 18, 413

Khatri R., Sunyaev R. A., 2013, Journal of Cosmology and Astroparticle Physics, 6, 26

Kibble T. W. B., 1980, Physics Reports, 67, 183

Kibble T. W. B., 1982, Acta Physica Polonica B, 13, 723

Kogut A., Abitbol M. H., Chluba J., Delabrouille J., Fixsen D., Hill J. C., Patil S. P., Rotti A., 2019, in BAAS, Vol. 51, p. 113

Kogut A., Chluba J., Fixsen D. J., Meyer S., Spergel D., 2016, in Proc.SPIE, Vol. 9904, SPIE Conference Series, p. 99040W

Kogut A., Fixsen D., Chuss D., Dotson J., Dwek E., et al., 2011, JCAP, 1107, 025

Kuroyanagi S., Nakayama K., Yokoyama J., 2015, PTEP, 2015, 013E02

Kuroyanagi S., Takahashi T., Yokoyama S., 2020, arXiv e-prints, arXiv:2011.03323

Lyth D. H., Riotto A., 1999, Phys. Rept., 314, 1

Machado C. S., Ratzinger W., Schwaller P., Stefanek B. A., 2019, JHEP, 01, 053

Machado C. S., Ratzinger W., Schwaller P., Stefanek B. A., 2020, Physical Review D, 102, 075033

Marsh D. J. E., 2016, Phys. Rept., 643, 1

Mather J. C. et al., 1994, The Astrophysical Journal, 420, 439

Mukherjee S., Khatri R., Wandelt B. D., 2018, Journal of Cosmology and Astroparticle Physics, 2018, 045

Nakai Y., Suzuki M., Takahashi F., Yamada M., 2020, arXiv e-prints, arXiv:2009.09754

Ota A., Takahashi T., Tashiro H., Yamaguchi M., 2014, Journal of Cosmology and Astroparticle Physics, 10, 29

Perera B., et al., 2019, Mon. Not. Roy. Astron. Soc., 490, 4666

Planck Collaboration et al., 2018, ArXiv:1807.06209

PRISM Collaboration et al., 2014, Journal of Cosmology and Astroparticle Physics, 2, 6

Reitze D., et al., 2019, Bull. Am. Astron. Soc., 51, 035

Sesana A. et al., 2019, arXiv e-prints, arXiv:1908.11391

Smith T. L., Caldwell R., 2019, Phys. Rev. D, 100, 104055

Sunyaev R. A., Zeldovich Y. B., 1970, Astrophysics and Space Science, 9, 368

Sunyaev R. A., Zeldovich Ya. B., 1970, Astrophys. Space Sci., 7, 20

Tashiro H., Sabancilar E., Vachaspati T., 2013, Journal of Cosmology and Astroparticle Physics, 8, 35

Thorne B., Fujita T., Hazumi M., Katayama N., Komatsu E., Shiraishi M., 2018, Phys. Rev. D, 97, 043506

Vilenkin A., Shellard E. P. S., 1994, Cosmic strings and other topological defects

Watanabe Y., Komatsu E., 2006, Phys. Rev. D, 73, 123515

Weinberg S., 2004, Physical Review D, 69, 023503

Weltman A., et al., 2020, Publ. Astron. Soc. Austral., 37, e002

Zeldovich Ya. B., Sunyaev R. A., 1969, Astrophys. Space Sci., 4, 301

5

Paper II: Clarifying transfer function approximations for the large-scale gravitational wave background in ΛCDM

This paper offers a follow up to the first paper (chapter 4 of this thesis). We revisit and model the transfer functions for tensor perturbations in the primordial Universe (see chapter 3), making use of the Physics described in Watanabe & Komatsu (2006) and Saikawa & Shirai (2018).

In performing the modelling from the ground up we are able to independently verify the results of Weinberg (2004); Dicus & Repko (2005), and demonstrate some potential mistakes which would easily go unnoticed in the literature. Firstly we find that analytic approximations based on a Universe only containing matter and radiation will only work using a corresponding modified current age of the Universe. Secondly we show that correctly accounting for the density of relativistic particles (as compared to only radiation) also provides a noticable correction. These previous two effects can mask eachother, and thus make it difficult to fully reproduce some results in the literature.

We also use the numerical treatment to verify the effects that late time acceleration (see chapter 1) would have on the background of gravitational waves, which amount mostly to phase changes in the low frequency part of the spectrum.

arXiv:2107.13351v1 [astro-ph.CO] 28 Jul 2021

Clarifying transfer function approximations for the large-scale gravitational wave background in ΛCDM

Thomas Kite[1]⋆, Jens Chluba[1]† Andrea Ravenni[1]‡, and Subodh P. Patil[2]§
[1]*Jodrell Bank Centre for Astrophysics, School of Physics and Astronomy, The University of Manchester, Manchester, M13 9PL, U.K.*
[2]*Instituut-Lorentz for Theoretical Physics, Leiden University, 2333 CA Leiden, The Netherlands*

Accepted 2020 –. Received 2020 –

ABSTRACT

The primordial gravitational wave background (GWB) offers an exciting future avenue of discovery for new physics. Its information content encodes multiple eras in the early Universe's history, corresponding to many orders of magnitude in frequency and physical scale to be measured today. By numerically solving for the GW transfer functions we provide simple yet accurate formulas describing the average power of the large-scale energy spectrum of the GWB for arbitrary primordial tensor power spectra. In doing so we can pedagogically explain and clarify previous GWB literature, highlight the important cosmological parameters of various GWB features, and reveal multiple ways in which cancelling conceptual errors can give deceptively accurate results. The scales considered here are particularly important for CMB probes of the GWB, via *B*-modes and spectral distortions. In particular, we carefully study the effects of both neutrino damping, and the precise nature of the transition between the radiation-dominated (RD) and matter-dominated (MD) eras. A byproduct of numerically solving the problem is the ability to study the robustness of common approximations in the literature. Specifically, we show that a numerical treatment is especially important around the RD–MD transition, and for a brief moment of history where neutrino damping occurs during MD. In passing we also discuss the effects of late acceleration caused by dark energy – showing that this can be neglected in most practical GWB applications – and the effects of changing relativistic degrees of freedom on the GWB at very small-scales.

Key words: cosmology: theory — gravitational waves —

1 INTRODUCTION

The detection of the first gravitational wave (GW) (LIGO Scientific Collaboration & Virgo Collaboration 2016) opened a door to a novel way of studying the Universe. Decades of studying the light arriving from the cosmos has provided us with modern precision cosmology as we know it, and with some poetic license, we are now able to *hear* the Universe as well as *see* it.

The excitement of this prospect has led to a suite of new upcoming probes (either proposed or under construction) which will listen for GWs in different frequency bands (see Campeti et al. 2021, for review). From lowest to highest frequency GWs we have CMB *B*-mode measurements (Ade et al. 2018; Aghanim et al. 2020), spectral distortion measurements (Kite et al. 2020), pulsar timing array measurements (Perera et al. 2019; Alam et al. 2020), and finally direct detection using interferometry (Abbott et al. 2020b,a). Through a combination of all these probes we can construct a comprehensive picture of the symphony of GWs in the Universe, and refine our understanding of fundamental physics in the process.

In this work, we focus on primordial origins of GWs rather than astrophysical sources. Our study therefore relates to searches for a stochastic gravitational wave background (GWB) rather than single isolated events. The exact physics that will be revealed through studying this background is broad and diverse (see Caprini & Figueroa 2018, for review).

The goal of this paper is then twofold: firstly to pedagogically introduce the physics of the GWB to clarify other literature, revealing potential pitfalls in the analytic modelling, and secondly to provide a simple yet accurate analytic description for the mapping between the present-day large-scale GWB energy spectrum and the corresponding primordial tensor power. The latter allows our results to be applied to general inflationary models, making this work particularly relevant to the interpretation of *B*-mode and spectral distortion searches for new physics.

The mapping from underlying physical model to present-day observations requires a detailed understanding of the GW transfer function, for which various solutions have been considered (e.g., Watanabe & Komatsu 2006; Dicus & Repko 2005; Caprini & Figueroa 2018). We expand upon this literature with a numerical treatment of the GWB which accounts for the nuanced cosmological expansion through radiation-dominated (RD) and matter-dominated (MD) eras, the late time accelerated expansion from dark energy (DE) and the non-negligible damping from free-streaming neutrinos. This allows us to give the promised simple fits for the average large-scale GWB energy spectrum in a number of fiducial cosmological scenarios.

Accurately accounting for the transition between RD and MD eras is especially important in calculations of the transfer function

---

⋆ E-mail: thomas.kite@manchester.ac.uk
† E-mail: jens.chluba@manchester.ac.uk
‡ E-mail: andrea.ravenni@manchester.ac.uk
§ E-mail: patil@lorentz.leidenuniv.nl

© 0000 The Authors

for non-standard thermal histories, such as those with epochs of early matter domination frequently encountered in a variety of phenomenological extensions of the standard cosmology (Acharya et al. 2008, 2019), or for scenarios where the primordial GW spectrum is significantly enhanced or modulated, relevant, for instance, in scenarios of primordial black hole formation (Ballesteros et al. 2020; Bhattacharya et al. 2021; Green & Kavanagh 2021; Arbey et al. 2021). We will discuss how the results of this paper can also be straightforwardly extended to such applications.

This paper is organised as follows: in Sect. 2 we qualitatively review the broad range of fundamental physics imprinted on the GWB. This will aid the reader in understanding the more quantitative approach in Sect. 3, where we analytically solve the equation governing the evolution of GWs in limiting cases. These solutions, although previously considered, will serve to clarify some confusion in the literature about their application. The numerical method is explained and results shown in Sect. 4, focusing on the reliability of the analytic results previously found. One region of parameter space not captured well by existing approximations is the MD–RD transition, which is important for CMB scale probes. Hence in this section we provide simple fits for the large-scale GWB, providing an alternative to the usual analytic approximations. More general features of the GWB are discussed in Sect. 5, where we demonstrate the principal cosmological dependence of neutrino damping and the main effects of late time acceleration on the GWB. For completeness we include some discussion of changes in the relativistic degrees of freedom and their relevance to SD constraints on GW backgrounds. We point out in this section how a combination of the simple fits and pretabulated data on relativistic degrees of freedom can accurately model the spectrum to within $\sim 5\%$ on all scales. Finally we summarise and conclude in Sect. 6.

2 PHYSICS CONTENT OF THE GWB

The study of cosmological perturbation theory explains the evolution of perturbations on the otherwise smooth expanding FLRW background, and is the foundation for much of modern cosmology (Ma & Bertschinger 1995). For detailed explanation and derivations with details about GWs see Weinberg (2008), but we summarise the essential steps here. Perturbatively small terms are added to both the metric $g_{\mu\nu}$ and the stress energy tensor $T_{\mu\nu}$, which can then be equated through Einstein's field equations. Three fundamental types of perturbations emerge from this calculation: scalars, transverse vectors and spatial transverse traceless tensors. The latter are what we also understand as GWs. These waves couple to the corresponding spatial transverse traceless tensor component within $T_{\mu\nu}$, the anisotropic stress of the medium, Π, which provides a source term that can damp the GWB.

This last point is quite important, as typically speaking the particle species in the primordial plasma do not carry considerable anisotropic stresses: tightly-coupled fluids rapidly isotropize and are dominated by their densities and velocities, after which comes a period of free streaming dominated solely by velocity[1]. Only a brief intermediate phase therefore leads to a non-negligible anisotropic stress that can interact with and damp the GWB. The dominant damping effects therefore arise from the GWs themselves sourcing the anisotropic stress in the medium, which will lead to an integro-differential equation that we solve numerically.

A subdominant contribution to the damping is added by the cosmic photon field. At early times the photon fluid inherits enough energy from the GWB to produce a noteworthy distortion to the blackbody spectrum (Chluba et al. 2015), but with no discernible effect on the GWB. The GWB scales most affected by photons are $k \simeq 10^{-2}\,\mathrm{Mpc}^{-1}$, amounting to a 14% reduction in the amplitude squared according to the work of Saikawa & Shirai (2018). However, we note that at these scales it is both possible and necessary to model the photon decoupling with the full Boltzmann equation, rather than using a modified version of the damping term [e.g., Eq. (17a) below], which contains several simplifying assumptions. The damping effect of photons will not significantly change the results of this paper, and a full detailed treatment is left to future work.

The neutrino, on the other hand, has a considerable damping effect over a large set of scales. Previous studies show that the neutrino field will damp the GWB amplitude squared by $\simeq 35.6\%$ (Weinberg 2004; Dicus & Repko 2005) at scales $k \gtrsim 1/500\,\mathrm{Mpc}^{-1}$. The damping effect arising from neutrinos will be investigated below, verifying and generalising on these previous studies. We note that it is conceivable to treat the neutrino field with the same level of sophistication as the photon field: understanding how inherited energy from the GWB will distort the otherwise thermal distribution of neutrino momenta, and modelling a gradual decoupling of the particles through full Boltzmann hierarchies. However, also this program is beyond the scope of this paper.

The bottom line then is that within the standard thermal history of the Universe, the GWB is mostly free from the surrounding plasma, only receiving small predictable damping effects from free streaming neutrinos. The rest of the information encoded in the GWB therefore comes from the state of the Universe at the time of horizon crossing for each frequency, after which simple propagation occurs. This is, in fact, the double-edged sword of GW cosmology: a feeble interaction that simultaneously makes a clean and powerful probe of almost the entirety of cosmological history, but which also makes for an incredibly difficult detection at present time. A detection is a sufficiently monumental task that glimpsing the GWB has become the aspiration of many scientific teams, with a diverse set of probes.

One important state of the Universe's history cleanly imprinted as a GWB feature is the precise moment that relativistic particle species no longer dominate the universal expansion, giving way to a matter dominated era. Since GWs have a different evolution in each of the eras, there is a predictable change in shape of the energy spectrum (see Sect. 4.2). One goal of this paper is to elucidate this transition in order to facilitate comparison between early and late Universe probes of the GWB.

To model the moment of this transition it is important to cleanly separate the cosmic inventory into relativistic and non-relativistic particles. This usually equates to distinguishing massive and massless species, but some subtleties arise when considering neutrinos. We now know from data on neutrino oscillations (Fukuda et al. 1998; Ahmad et al. 2001, 2002) to expect massive neutrinos, albeit with masses limited to sub-eV scales (Planck Collaboration et al. 2018b; Aker et al. 2021). The concordance model in Cosmology therefore still treats these as massless entities in most applications. This is often sufficient since the sum of neutrino masses is predicted to be sufficiently small that the early-universe dynamics will resemble that of massless particles, even if at least two of the neutrino species must be non-relativistic today (Lesgourgues & Pastor 2006).

In this paper, we therefore carefully distinguish the photon energy density, $\Omega_\gamma = 5.42 \times 10^{-5}\,[T_0/2.7255\,\mathrm{K}]^4\,[h/0.675]^{-2}$, from the total relativistic energy density

---

[1] For a more general analysis that interpolates between the kinetic and hydrodynamic regimes, incorporating ambient matter interactions, see for instance (Baym et al. 2017; Flauger & Weinberg 2018; Mirón-Granese 2020; Zarei et al. 2021).

$$\Omega_{\rm rel} = \Omega_\gamma + \Omega_\nu = \left(1 + N_{\rm eff}\left[\frac{7}{8}\right]\left[\frac{4}{11}\right]^{4/3}\right)\Omega_\gamma, \tag{1}$$

which includes the neutrino energy density Ω_ν. The number of relativistic degrees of freedom, $N_{\rm eff}$, parameterizes the extra massless degrees of freedom relative to the photons. The factor of $7/8$ arises due to the differences in particle statistics (Bose-Einstein or Fermi-Dirac), while the factor $(4/11)^{4/3}$ relates to the energy release during electron-positron annihilation. In this paper we assume the standard model expectation value of $N_{\rm eff} = 3.046$ (Mangano et al. 2005; de Salas & Pastor 2016), which in turn gives $\Omega_{\rm rel} = 9.18 \times 10^{-5}$ today. This distinction between the photon field and the full relativistic cosmic inventory has been ambiguous or neglected in some literature, leading to additional confusion around the exact moment of RD–MD transition (e.g. see discussion in sect. 5.2 in Caprini & Figueroa 2018). As previously mentioned, resolving this disparity is important for accurate comparison between the largest scale CMB B-Modes and spectral distortion measurements, and constitute one driving motivation for this work.

One more energy component needs to be included to complete the cosmic inventory: the cosmological constant or dark energy[2] Ω_Λ. Despite being the dominant form of energy today, it makes up a tiny fraction of the Universe's content at primordial times. The expected effect of this component is only small changes on the largest physical scales, which can be verified numerically (Sect. 5.2). A more notable difference from the late-time acceleration is the change in the age of the Universe, which complicates the application of analytic solutions, as we clarify here.

The physics discussed thus far is all needed to accurately model the GWB down to scales of $k \simeq 10^3\,{\rm Mpc}^{-1}$. Beyond these scales the spectral features arise from changes in the number of relativistic degrees of freedom, g_*, as originally discussed in Watanabe & Komatsu (2006), generalised by Boyle & Steinhardt (2008), and recently solved to high precision by Saikawa & Shirai (2018). These changes in the energy budget, arising from the cooling effect of the universal expansion, cause small temporary changes in the expansion rate, which is imprinted on the GWB from the moment of horizon crossing. We will briefly discuss the importance of these effect on spectral distortion constraints, leaving the details of the physics to the aforementioned papers.

3 ANALYTIC GW SOLUTIONS

The equation of motion governing the evolution of a GW, derived from cosmological perturbation theory, is given by (Weinberg 2004; Watanabe & Komatsu 2006; Boyle & Steinhardt 2008)

$$\partial_\eta^2 h_k^\lambda + 2\frac{a'}{a}\partial_\eta h_k^\lambda + k^2 h_k^\lambda = 16\pi G a^2 \Pi^\lambda, \tag{2}$$

where $h_k^\lambda(\eta)$ is the amplitude of the gravitational wave at wavenumber k for each polarization $\lambda = +, \times$, and $\Pi^\lambda(k,\eta)$ is the anisotropic stress of the surrounding primordial plasma, both as a function of wavenumber k and conformal time η. Primes denote derivatives respect to conformal time, but we keep some explicit derivatives for clarity later where we will change coordinates. The amplitude of a physical GW can be written as the product of a transfer function with some initial amplitude $h_k^\lambda(\eta) = h_k^{\lambda,{\rm prim}}\mathcal{T}_{\rm GW}(k,\eta)$, and as such we have $\mathcal{T}_{\rm GW}(k,0) = 1$. This decomposition of transfer function and initial condition helpfully separates the statistical from the deterministic, as well as distinguishing the inflationary from the post-reheating dynamics.

A primary goal of this paper is to give simple yet precise estimates for the energy density of the GWB, which measured relative to the critical density is given by

$$\Omega_{\rm GW}(k) = \frac{\rho_{\rm GW}}{\rho_{\rm c}}(k) = \frac{\mathcal{P}_T(k)}{12a^2H^2}[\mathcal{T}'_{\rm GW}(k)]^2. \tag{3}$$

Here, the primordial tensor power spectrum

$$\mathcal{P}_T(k) = \frac{2k^3}{2\pi^2}\sum_\lambda \langle |h_k^{\lambda,{\rm prim}}|^2\rangle \tag{4}$$

encodes the statistical properties of the initial conditions via an ensemble average[3]. For many applications the energy density is the essential quantity one needs to know, since any experiment measuring the GWB is sensitive to its energy density at a given time and scale/frequency. It is clear from Eq. (3) that fundamental link between the primordial $\mathcal{P}_T$ and $\Omega_{\rm GW}$ at any other time is the transfer function $\mathcal{T}_{\rm GW}$, which we study in detail next.

3.1 Transfer function

As previously discussed, a key feature in the GWB is a distinctive *bend* on physical scales corresponding to the transition between the radiation-dominated[4] (RD) and matter-dominated (MD) eras of the Universe's history. To understand this effect it is instructive to first ignore both the contribution of DE and the effects of damping - the former being negligible and the latter being an unnecessary complication to describe the physics of the transition. Solving the Friedman Equations in this limit we have

$$\eta = 2\frac{\sqrt{a\Omega_{\rm m} + \Omega_{\rm rel}} - \sqrt{\Omega_{\rm rel}}}{H_0\Omega_{\rm m}}, \tag{5a}$$

$$a = \frac{1}{4}\eta^2 H_0^2\Omega_{\rm m} + \eta H_0\sqrt{\Omega_{\rm rel}}, \tag{5b}$$

$$\frac{a'}{a} = aH = aH_0\sqrt{\Omega_{\rm m}a^{-3} + \Omega_{\rm rel}a^{-4}}. \tag{5c}$$

Using these expressions one can find

$$\frac{a'}{a} = \frac{1}{\eta} + \frac{1}{\eta + \eta_*} = \frac{1}{\eta_*}\left(\frac{1}{\xi} + \frac{1}{1+\xi}\right), \tag{6a}$$

$$\eta_* = 1/k_* = 4\sqrt{\Omega_{\rm rel}}/H_0\Omega_{\rm m}. \tag{6b}$$

The characteristic time-scale defined here is $\eta_* = 540.44\,{\rm Mpc}$ for up-to-date cosmological parameters from Planck Collaboration et al. (2018a). With this time-scale, the dimensionless quantities $\xi = \eta/\eta_*$ and $\kappa = k/k_* = k\eta_*$ naturally emerge. Using these variables is advantageous for various reasons, but most notably it adds a degree of invariance in considering different cosmologies. Note the commonly appearing term $\kappa\xi = k\eta$, which helps in matching to common approximations in the literature. Another common time-scale for RD–MD equality is $a_{\rm eq} = \Omega_{\rm rel}/\Omega_{\rm m}$, defined simply as the time in which energy densities of the respective components matched[5].

---

[2] For the purposes of this paper, Ω_Λ will be referred to as dark energy and cosmological constant interchangeably – only dark energy with $w = -1$ is considered.

[3] We have followed the convention of Watanabe & Komatsu (2006) and Saikawa & Shirai (2018), which can be expressed in terms of other conventions by noting the normalisation of polarisation tensors in the latter reference, between Eq. (2.4) and Eq. (2.5).

[4] We remind the reader that despite the misnomer we include relativistic neutrinos here.

By incorporating this change of variables to the differential equation we find an elegant form

$$\partial_\xi^2 \mathcal{T}_{\rm GW} + 2\left(\frac{1}{\xi} + \frac{1}{1+\xi}\right)\partial_\xi \mathcal{T}_{\rm GW} + \kappa^2 \mathcal{T}_{\rm GW} \approx 0. \tag{7}$$

The characteristic time-scale used here can be further motivated by noticing that it is the time that balances the two contributions to Eq. (5b), showing it is closely related to the balance of matter and radiation.

Using Eq. (7) it is possible to study the evolution of GWs far into both RD ($\xi \ll 1$) and MD ($\xi \gg 1$). In each limiting case we obtain

$$2\left(\frac{1}{\xi} + \frac{1}{1+\xi}\right) \longrightarrow \begin{cases} 2/\xi & \text{for RD} \\ 4/\xi & \text{for MD} \end{cases}, \tag{8}$$

which offer simple solutions to Eq. (7) in terms of spherical Bessel functions which we summarise here:

$$\mathcal{T}_{\rm GW}^{\rm RD} = A\, j_0(\kappa\xi) - B y_0(\kappa\xi), \tag{9a}$$

$$\mathcal{T}_{\rm GW}^{\rm MD} = \frac{3}{\kappa\xi}\left[C j_1(\kappa\xi) - D y_1(\kappa\xi)\right], \tag{9b}$$

with derivatives

$$\mathcal{T}_{\rm GW}^{\prime\,\rm RD} = -k\left[A j_1(\kappa\xi) - B y_1(\kappa\xi)\right], \tag{10a}$$

$$\mathcal{T}_{\rm GW}^{\prime\,\rm MD} = -\frac{3k}{\kappa\xi}\left[C j_2(\kappa\xi) - D y_2(\kappa\xi)\right]. \tag{10b}$$

Here A, B, C and D are constants determined from initial conditions and matching conditions which we discuss below. Note that the derivatives here are still with respect to conformal time, which yields factors of k. The terms involving spherical Bessel functions of the first kind, j_n, are constant at early times, and have been scaled here such that $A = C = 1$ gives an early time normalisation to unity. Spherical Bessel functions of the second kind, y_n, are the decaying modes.

The solutions given above are each valid deep into each regime, but we have yet to discuss the transition between them. Note first of all that MD scales ($\kappa \ll 1$) simply stay constant in the RD era, since for those modes we have $\kappa\xi \ll 1$. On the contrary we must be careful with the RD scales ($\kappa \gg 1$) during the MD era, since these modes have already had time to evolve and decay by that time. An approximation for this matching process is performed by Watanabe & Komatsu (2006) (henceforth WK06), where by assuming an instantaneous transition one can solve

$$\left.\mathcal{T}_{\rm GW}^{\rm RD}\right|_{\xi=1} = \left.\mathcal{T}_{\rm GW}^{\rm MD}\right|_{\xi=1}, \tag{11a}$$

$$\left.\mathcal{T}_{\rm GW}^{\prime\,\rm RD}\right|_{\xi=1} = \left.\mathcal{T}_{\rm GW}^{\prime\,\rm MD}\right|_{\xi=1}, \tag{11b}$$

which gives a functional form to the constants previously defined:

$$A = 1, \tag{12a}$$

$$B = 0, \tag{12b}$$

$$C(\kappa) = \frac{1}{2} - \frac{\cos(2\kappa)}{6} + \frac{\sin(2\kappa)}{3\kappa}, \tag{12c}$$

$$D(\kappa) = -\frac{1}{3\kappa} + \frac{\kappa}{3} + \frac{\cos(2\kappa)}{3\kappa} + \frac{\sin(2\kappa)}{6}. \tag{12d}$$

This matches the equations given by WK06 once accounting for different variable conventions. A noteworthy difference in convention is the lack a step function that enforces $C \to 1$ and $D \to 0$ for $\kappa \ll 1$, which is already the natural tendency of the functions as they are written here. Notice that the constant mode solution for the RD scales will excite a decaying mode in the MD era, and thus D cannot be ignored even if B has been.

3.2 Energy spectrum

By using the analytic forms derived in Sect. 3.1 the expected limits of the energy spectrum Eq. (3) can be derived. Recalling that we are only interested in the power spectrum normalised energy density as seen today ($\eta_0 \gg \eta_*$), we can take the limit in the MD era:

$$\begin{aligned} \left.\frac{\Omega_{\rm GW}}{\mathcal{P}_T}\right|_{\xi_0} &= \frac{1}{12H_0^2}\left[\mathcal{T}_{\rm GW}^{\prime\,\rm MD}\right]^2 \\ &= \frac{1}{12H_0^2}\frac{9}{\eta_*^2\xi_0^2}\left[C j_2(\kappa\xi_0) - D y_2(\kappa\xi_0)\right]^2, \end{aligned} \tag{13}$$

from which the high and low κ limits can be derived. In both limits however we note that any realistic probe of the GWB will have sensitivity on scales much smaller than those crossing horizon in recent times ($k \gg 1/\eta_0$). This statement leads to $\kappa\xi_0 \gg 1$ expansions, for which we have (Watanabe & Komatsu 2006)

$$j_n(x) \approx \frac{\sin(x - n\pi/2)}{x} \quad \text{for } x \gg 1, \tag{14a}$$

$$\langle j_n(x)^2 \rangle \approx \left(\frac{1}{2}\right)\frac{1}{x^2} \quad \text{for } x \gg 1, \tag{14b}$$

where angle brackets indicate averages over an oscillation, leading to an explicit[6] factor of $1/2$.

For fixed ξ_0 and *large* κ, the dominant term in the expansion of Eq. (14) will have a linear term $\kappa/3 \subset D(k)$ combined with $-\cos(\kappa\xi_0)/\kappa\xi_0 \subset y_2(\kappa\xi_0)$. This gives a flat (albeit oscillating) spectrum to high frequencies:

$$\begin{aligned} \left\langle \left.\frac{\Omega_{\rm GW}}{\mathcal{P}_T}\right|_{\xi_0} \right\rangle &\overset{\substack{\kappa\gg1 \\ \downarrow}}{\approx} \frac{9}{12H_0^2\eta_*^2\xi_0^2}\left\langle \left[-\frac{\kappa}{3}\frac{\cos(\kappa\xi_0)}{\kappa\xi_0}\right]^2 \right\rangle \\ &\overset{\substack{\kappa\xi_0\gg1 \\ \downarrow}}{\approx} \left(\frac{1}{2}\right)\frac{\eta_*^2}{12H_0^2\eta_0^4} \\ &= \left(\frac{1}{2}\right)\frac{\Omega_{\rm rel}}{12}. \end{aligned} \tag{15}$$

Note however in the final equality we have used a value of η_0 derived from Eq. (5b), by setting $a_0 = 1$. This may appear problematic, since dark energy dominates the expansion from $a \gtrsim 3/4$, and thus changes the age of the Universe. The analytic approximations derived here however were derived explicitly in a Universe without DE, and should not be used in conjunction with DE-modified values of η_0. This *cancellation of errors* is vindicated by the numerical solutions (see Sect. 5.2).

To investigate the behaviour at *low* κ, we start with $C(\kappa) \to 1$, $D(\kappa) \to 0$, and again apply the subhorizon condition $\kappa\xi_0 \gg 1$. This suggests the dominant term being $-\sin(\kappa\xi)/\kappa\xi \subset j_2(\kappa\xi)$. A similar

---

[5] It is often unclear which time-scale an author uses, and as such we will keep a strict convention here. The $\mathcal{T}_{\rm GW}$ approximations by Watanabe & Komatsu (2006), which we discuss shortly, give the correct limiting cases using η_* as defined in this work.

[6] For clarity we will keep this convention of explicit $1/2$ throughout the paper.

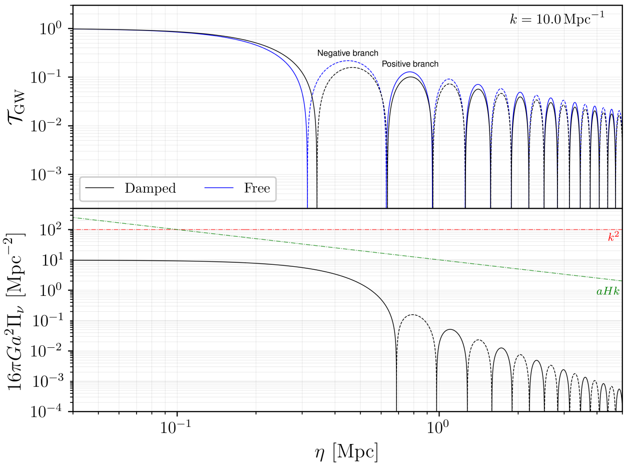


Figure 1. A figure showing an example of the neutrino damping term and its effect on the transfer function at $k = 10\,\mathrm{Mpc}^{-1}$. The top panel reveals small phase shifts accompany a drop in amplitude of $\mathcal{T}_{\mathrm{GW}}$ around the time of horizon crossing. Dashed lines indicate negative branches of the oscillating function. The lower panel shows the damping term $16\pi Ga^2\Pi_\nu$. Dotted lines show other terms in the differential equation associated with the wavenumber and Hubble expansion.

calculation to above gives

$$\left\langle \frac{\Omega_{\mathrm{GW}}}{\mathcal{P}_T}\bigg|_{\xi_0} \right\rangle \overset{\kappa\ll 1}{\downarrow \atop \approx} \frac{1}{12H_0^2}\frac{9}{\eta_*^2\xi_0^2}\left\langle \left[-\frac{\sin(\kappa\xi_0)}{\kappa\xi_0}\right]^2 \right\rangle \overset{\kappa\xi_0\gg 1}{\downarrow \atop \approx} \left(\frac{1}{2}\right)\frac{\Omega_{\mathrm{rel}}}{12}\frac{9}{\kappa^2}. \tag{16}$$

These results will be used in Sect. 4.2 to choose a functional form for an envelope fit to the data, and in turn verify the accuracy of the numerical calculations.

3.3 Anisotropic stress

We previously discussed that some particles will contribute to the anisotropic stress of the medium, and constitute damping terms to the GW solution. These stresses are excited by the propagation of the GW itself, and hence makes Eq. (2) an integro-differential equation, requiring a more careful treatment than the previous section. We do not derive any analytic solutions here, but instead quote the results of Dicus & Repko (2005) (henceforth DR04). We give the form of the damping integral here both for completeness, and to motivate a numerical approach to solving the problem, as described in Sect. 4.

Explicitly evaluating the RHS of Eq. (2) for the case of neutrinos gives (Weinberg 2004)

$$16\pi Ga^2\Pi_\nu^\lambda = -24 f_\nu \left(\frac{a'}{a}\right)^2 \int_{\eta_\nu}^{\eta} K(k[\eta-\bar\eta])\,\mathcal{T}'_{\mathrm{GW}}(\bar\eta)\,h_k^{\lambda,\mathrm{prim}}\,\mathrm{d}\bar\eta, \tag{17a}$$

$$f_\nu = \frac{\Omega_\nu}{\Omega_\gamma+\Omega_\nu}\frac{1}{1+a/a_{\mathrm{eq}}} = \frac{f_{\nu,0}}{1+a/a_{\mathrm{eq}}}, \tag{17b}$$

$$K(x) = \frac{1}{16}\int_{-1}^{1}(1-s^2)^2 \mathrm{e}^{isx}\mathrm{d}s = \frac{j_2(x)}{x^2}$$
$$= \frac{3\sin(x)}{x^5} - \frac{3\cos(x)}{x^4} - \frac{\sin(x)}{x^3}, \tag{17c}$$

where η_ν is the time at which neutrinos decouple, corresponding to a temperature of $\simeq 2$ MeV (e.g., Jeong et al. 2014).

An example of this damping term is shown in Fig. 1, where it is seen that at the time of horizon crossing there is a significant damping of the wave followed by a period of regular propagation, albeit at a lower overall amplitude. The dotted lines show approximate amplitudes for other terms in Eq. (2), revealing that the damping is subdominant, and comparable only at horizon crossing.

We note in passing that more general particle interactions in the collision time approximation contribute an additional exponential suppression inside the integrand of Eq. (17a) of the form $\exp\left[-\int_{\bar\eta}^{\eta}\frac{d\eta'}{\tau_c(\eta')}\right]$, where τ_c is the average time between particle collisions (Baym et al. 2017), making manifest that tightly coupled particles rapidly isotropize and suppress anisotropic stresses, whereas free streaming particles, for which $\tau_c \to \infty$ reduces to Eq. (17a). We also note that Saikawa & Shirai (2018) use a modified expression for the neutrino energy fraction f_ν which includes energy inherited from e^+e^- annihilation. This leads to a slightly greater damping effect at scales of $k \sim 3\times10^4\,\mathrm{Mpc}^{-1}$, quickly adopting the same asymptotic limit as found in this paper (See Fig. 8).

Within the RD era, the damped transfer functions are given by DR04 in the form of a series sum of spherical bessel functions:

$$\mathcal{T}_{\mathrm{GW}}(k\eta) = \sum_{n=0}^{\infty} a_{2n}\,j_{2n}(k\eta). \tag{18}$$

Although in principle this sum has infinitely many terms, in practice only a few are needed. We will take this series with the 7 coefficients provided by DR04 as a benchmark in the RD era, but differences are expected as the universe becomes more matter dominated.

4 NUMERICAL SOLUTIONS

Evaluating the damped solution (see Sect. 3.3) has all the usual difficulties of an integro-differential equation: it involves an integral over the history of the GW's own velocity, and it cannot be easily pretabulated since the integrand depends on the upper limit of the integral itself. In this work we use an iterative method to achieve the solution to within some desired accuracy: the first iteration of the method assumes no damping [therefore solving Eq. (7)] to achieve an initial *guess* $\mathcal{T}_{\mathrm{GW}}^{(0)}$. Each subsequent iteration calculates $\mathcal{T}_{\mathrm{GW}}^{(N)}$ by inserting $\mathcal{T}_{\mathrm{GW}}^{(N-1)}$ to the damping integral. This has the advantage of allowing the damping term to be precalculated for a series of values of η, and then interpolated ready to use in a new iteration of the ODE solution. This makes the solution tractable, even if still somewhat numerically expensive, involving $O(N)$ integrals for a total $O(N^2)$ algorithm.

The iteration process ends by some metric of convergence. Here we sum the squared residuals between consecutive solutions for $\mathcal{T}'^{(N)}_{\mathrm{GW}}$ and divide by the number of timesteps and the wavenumber k. The former division guarantees an intensive metric for convergence – independent of number of points considered – and the latter accounts for the derivatives being $\simeq k$ larger than the transfer functions[7] which are bounded $-1 \leqslant \mathcal{T}_{\mathrm{GW}} \leqslant 1$.

Depending on the chosen wavenumber and desired precision, the method takes $\simeq$ 5–10 iterations to reach a final converged solution. This takes just a few seconds after moderate optimisation, using ODE solvers in the anisotropy module of `CosmoTherm` (Chluba & Sunyaev 2012). This made it possible, with parallelisation, to

---

[7] Alternatively residuals between undifferentiated $\mathcal{T}_{\mathrm{GW}}$ can be considered, but derivatives are already stored in memory for the damping integral, hence this approach constitutes a memory saving.

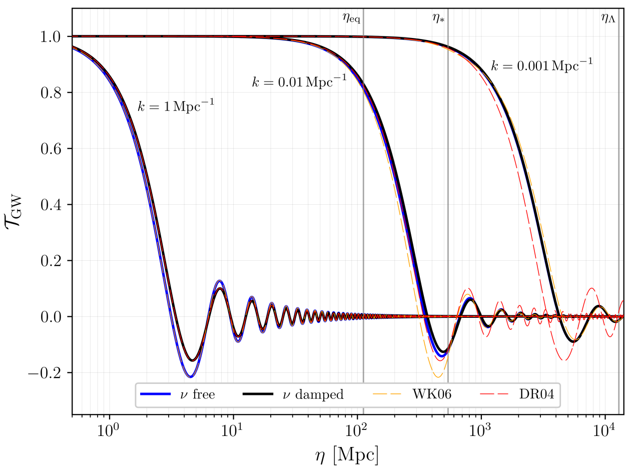


Figure 2. A graph showing examples of $\mathcal{T}_{\rm GW}$ for $k/{\rm Mpc}^{-1} \in [1, 0.01, 0.001]$. Cases with neutrino damping (black) and without (blue) are shown. The former is approximated by WK06 (orange) and the latter by DR04 (red). It can be seen that the free function is well approximated in each era, but not in the RD–MD transition. The damped function was approximated only in the RD era, hence showing large discrepancies at late times. It should be noted, however, that shortly after the transition the damping becomes negligible.

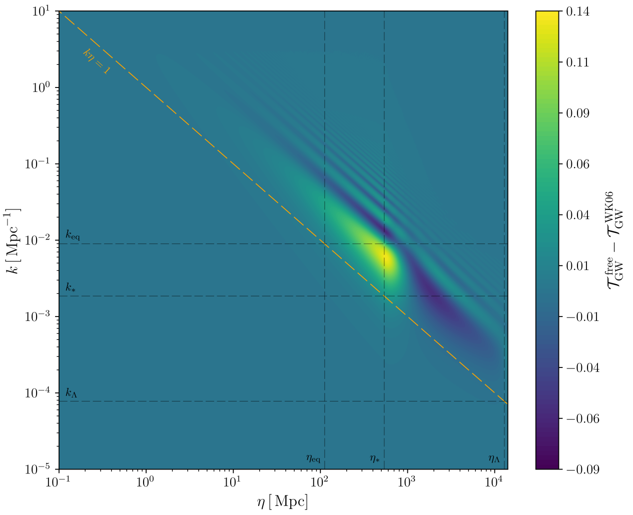
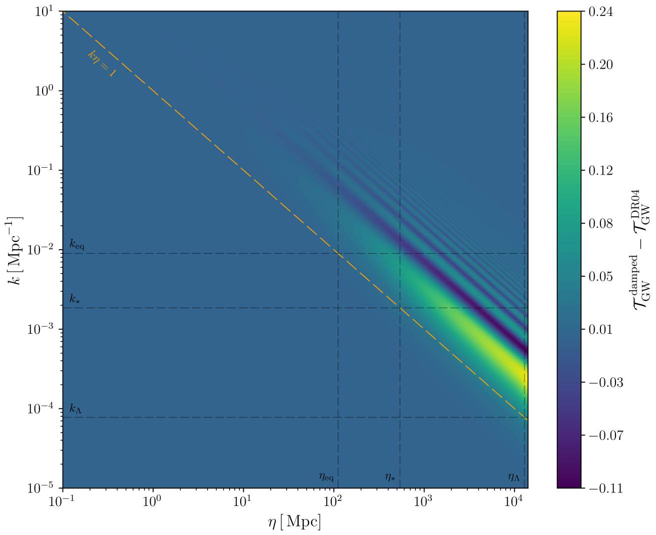


Figure 3. Contour plots showing the difference between this work's numerical solutions and approximations given by WK06, against both wavenumber k and conformal time η. An orange dashed line shows $k\eta = 1$. Gray dashed lines show RD–MD transition scales, $k_{\rm eq}$ and k_*, as defined in Sect. 3.1

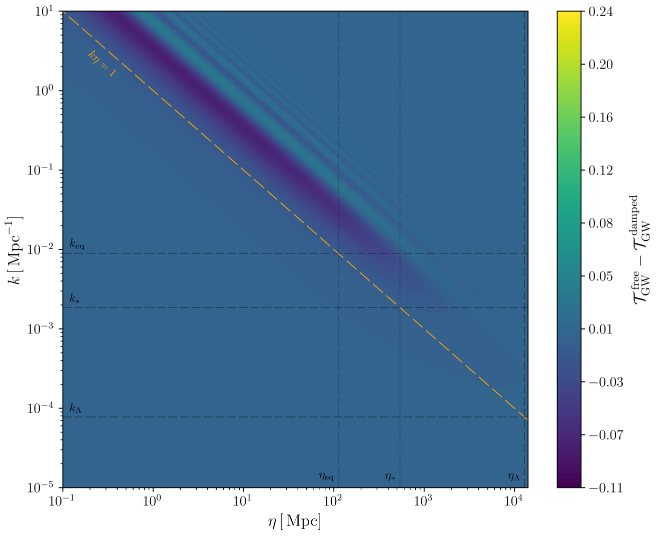
quickly solve the many tens or hundreds of thousands of k values needed to fit accurate envelopes to energy spectra.

4.1 Comparison to analytic results

To compare the analytic and numerical results we show a qualitative comparison [Fig. 2], and quantitative comparisons [Figs. 3,4]. The former illustrates that the waves usually differ more by an offset in phase than a difference in overall amplitude. With that in mind, we can properly interpret the contour plots, which reveal residuals oscillating throughout the parameter space. This suggests that integrated quantities across either time or wavenumber would be more accurate than these figures initially suggest.



(a) DR04 residuals



(b) Damping effects

Figure 4. Contour plots showing where the presence of damping is most important (top), and differences between this work's numerical solutions and approximations given by DR04 (bottom). This shows that the residuals in the right panel are mostly driven by the end of RD, with only a small intermediate phase showing both MD and damping behaviour simultaneously. An orange dashed line shows $k\eta = 1$. Gray dashed lines show RD–MD transition scales, $k_{\rm eq}$ and k_*, as defined in Sect. 3.1.

Deep in the RD era, we see an excellent agreement with both WK06 and DR04 as expected. Deep in the MD era on the other hand we see that we again have good agreement with WK06, albeit worse than before due to the matching conditions which essentially provide the MD initial conditions. Exactly at the transition is where the most discrepancy is seen, although the RD–MD transition is relatively short lived [Fig. 3]. The approximation by DR04 becomes progressively worse into the MD era, since their coefficients were derived assuming a RD Universe, accounting for the dominant part of the residuals [Fig. 4a]. We also provide Fig. 4b, which reveals

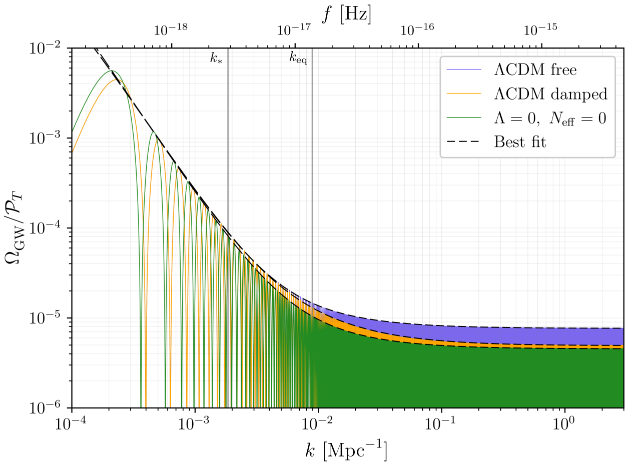


Figure 5. A graph showing $\Omega_{\rm GW}/\mathcal{P}_T$ across k as seen at η_0. Two distinct regions are discernible: a $1/k^2$ slope and a flat branch, corresponding to modes entering horizon during MD and RD, respectively. Spectra are shown for full ΛCDM with and without neutrino damping, and a simplified cosmology with $\Lambda = 0$, $N_{\rm eff} = 0$. The best-fit envelopes (i.e., twice the average power) for these cosmologies are shown with black lines, and given in Table 1.

how quickly the damping ceases once the MD era starts. Through a comparison of the two figures therefore we can see that most of the DR04 residuals are not from a poor modelling of damping, but simply from not capturing the MD dynamics.

In summary these comparisons highlight the robustness of the analytic approximations as well as the utility of the numerical solution by showing the latter can fill the gaps expected from former, but only in specific and brief regimes.

4.2 Simple fits for the large-scale GWB

In this subsection, we will give approximate fits for the energy spectrum of the GWB as measured at η_0 derived from the fully numerical treatment. This approach is greatly facilitated by knowing a sensible functional form in which to package the results, as was discussed in Sect. 3.2. Expecting a spectrum that interpolates between κ^{-2} and κ^0 motivates a more general formula for approximating the numerical results, which simply includes more general powers to capture the subtleties of the RD–MD transition[8]. As well as including powers between −2 and 0 we also include an inverse cubic term[9], which specifically in the ν damped scenario helps with the sharper increase in the spectrum as damping ceases for low k:

$$\left\langle \frac{\Omega_{\rm GW}}{\mathcal{P}_T}\bigg|_{\xi_0} \right\rangle = \left(\frac{1}{2}\right) \mathcal{D}\, \frac{\Omega_{\rm rel}}{12} \left(1 + \alpha_1 \kappa^{-1} + \alpha_2 \kappa^{-3/2} + \alpha_3 \kappa^{-2} + \alpha_4 \kappa^{-3}\right), \tag{19}$$

where $\mathcal{D}$ is a coefficient to represent neutrino damping that we discuss more generally in Sect. 5.1. The quality of this fit can be seen in Fig. 5, with corresponding coefficients given in Table 1. For ease of comparison three fiducial Cosmologies have been chosen: the first two corresponding to a best-fit *Planck* 2018 Universe, with and without neutrino damping, while the third is a simplified cosmology neglecting both neutrinos and Ω_Λ. This third cosmology highlights how simultaneously neglecting the neutrino contribution to the energy budget and the consequent neutrino damping will coincidentally lead to almost correct results, departing from the full solution by only ∼ 6%. We hope by providing this fit it may be easier to diagnose oversights in the literature (a similar cancellation of mistakes with almost correct results will be discussed in Sect. 5.2).

By using the natural scale to define κ we yield simple values for the α coefficients, with some implicit degree of cosmology independence (Note the similarity between the first and third row in Table 1, once accounting for changes in η_*).

Although only shown to $k = 3\,{\rm Mpc}^{-1}$ here, that is sufficient to show the limit of the GWB envelope indeed tends to $\Omega_{\rm rel}/12$. This is important, as the amplitude of the spectrum can be extrapolated beyond this scale without full calculation: cosmological dependence is imprinted at the time of horizon crossing, and so evaluating a given solution to $k\eta \gtrsim 100$ with and without some physical effect allows one to extrapolate[10] the spectrum appropriately by multiplying the appropriate ratio of transfer functions by $\Omega_{\rm rel}/12$. We use this to model the effects of changing relativistic degrees of freedom in Sect. 5.3, which all occur at scales $k \gtrsim 10^3\,{\rm Mpc}^{-1}$.

5 ANALYSIS OF GWB FEATURES

In this section, we discuss some of the GWB features previously mentioned in closer detail, with specific focus on the consequences for SD constraints on the GWB. We start with aspects of neutrino damping, then cover the late dark energy domination and finish with a discussion of the early thermal history.

5.1 Neutrino damping

By comparing the solution to Eq. (2) with and without the damping integral, and ignoring phase shifts of the transfer function, we can define the damping factor $\mathcal{D}$ as the ratio of the amplitudes squared: $\left[\mathcal{T}_{\rm GW}^{\rm damped}\right]^2 \approx \mathcal{D}\left[\mathcal{T}_{\rm GW}^{\rm free}\right]^2$. This definition mirrors that introduced in Sect. 4.2, but can now be applied to a single wave of wavenumber k. This is useful since the damping will only affect a finite range of scales, essentially giving $\mathcal{D} \equiv \mathcal{D}(k)$.

Specifically we expect $\mathcal{D}$ to tend to unity both for large and small k – the former since the modes were subhorizon before neutrinos started free-streaming, and the latter because the energy density of neutrinos was too small to have considerable effects. The low-k shape of $\mathcal{D}$ is intrinsically linked with the MD–RD transition, a moment which was relatively recent in cosmological history. The exact low-k dependence of $\mathcal{D}$ therefore does not manifest clearly in the GWB as seen today [e.g. convergence of damped and undamped solutions would look differently in Fig. 5 if the Universe was older/younger]. Instead we turn our attention to the shape of the damping envelope for large k, which in contrast reveals itself clearly, as seen in Fig. 6 (for $\sigma = 0$, as introduced next). The figure shows an expected smooth transition from the previously discussed damping constant and unity, but with a plateau around $2 \times 10^3 \lesssim k/{\rm Mpc}^{-1} \lesssim 10^4$. This feature is associated with the energy introduced to the medium from electron-positron annihilation

---

[8] Other attempts included having $\kappa^{-1/2}$, and allowing for a general power κ^γ. The fit used in the main text was chosen through trial and error, showing better results with simpler coefficients than the other functional forms.

[9] This term was not included in previous work (Kite et al. 2020). The changes however are only a few percent, and outside the scales visible to μ distortions.

[10] Note that Saikawa & Shirai (2018) perform a similar extrapolation on the transfer functions themselves using the WKB approximation.

Cosmology	$\mathcal{D}$	$\Omega_{\rm rel}$	η^*	α_1	α_2	α_3	α_4
ΛCDM free	1	9.18×10^{-5}	540.44 Mpc	4.15	−4.55	11.08	−0.11
ΛCDM damped	0.642	9.18×10^{-5}	540.44 Mpc	8.06	−8.46	17.86	−0.20
$\Lambda=0$, $N_{\rm eff}=0$	1	5.43×10^{-5}	415.50 Mpc	4.17	−4.21	10.55	−0.01

Table 1. Coefficients to calculate $\Omega_{\rm GW}/\mathcal{P}_T$ from Eq. (19) for three fiducial Cosmologies.

(see Sect. 5.3), which prolongs the time at which the universe has $T\approx2\,$MeV, the temperature of neutrino decoupling.

The damping integral in Eq. (17a) assumed an instantaneous decoupling of the neutrinos, which leads to oscillations in the damping envelope, as noted in WK06. A more realistic scenario can be achieved by introducing a factor to the integrand which smoothly tends to 0 for $\eta<\eta_\nu$ and to unity for $\eta>\eta_\nu$, with some characteristic width σ governing the sharpness of transition:

$$\int_{\eta_\nu}^{\eta}(\cdots)\;\longrightarrow\;\int_0^{\eta}\left[\frac{1+\tanh\left(\frac{\eta-\eta_\nu}{\sigma}\right)}{2}\right](\cdots)\,. \qquad (20)$$

The effects of this σ are also shown in Fig. 6. Moving forward we adopt a fiducial value of $\sigma=0.2\eta_\nu$, which quickly converges to the correct limits without spurious oscillations. The curve can be approximately replicated by replacing the factor of $\mathcal{D}$ in the second row of Table 1 with

$$\mathcal{D}(\chi)\approx0.642+(1-0.642)\frac{(11.08\,\chi)^3-(10.78\,\chi)^2}{(11.08\,\chi)^3-(9.41\,\chi)^2+1}, \qquad (21)$$

where $\chi=k/3.5\times10^5$. This approximate curve is shown as a dashed line in Fig. 6. We will see in Sect. 5.3 that this can be used to replicate the entire GWB spectrum to arbitrarily large k.

Although this modified treatment of neutrino decoupling is by no means considered accurate, it highlights an important dependence of the precise shape of the damping envelope on the decoupling physics. As mentioned above, a more accurate treatment including the full decoupling, neutrino oscillations and possible neutrino spectral distortions should be considered, is, however, beyond the scope of this work.

The ceasing of damping effects at large k has not been included in the calculation of spectral distortion window functions (Chluba 2014; Kite et al. 2020), meaning these have been underestimated. With a full calculation we would see a boost in sensitivity of $\simeq36\%$ on scales $10^5\lesssim k/{\rm Mpc}^{-1}\lesssim10^8$. This nearly corresponds to a factor of two in the observing time, rendering this correction non-trivial. However, at $k\gtrsim10^8\,{\rm Mpc}^{-1}$, the effect of relativistic degrees of freedom become more important, almost exactly canceling this omission (again, see Sect. 5.3).

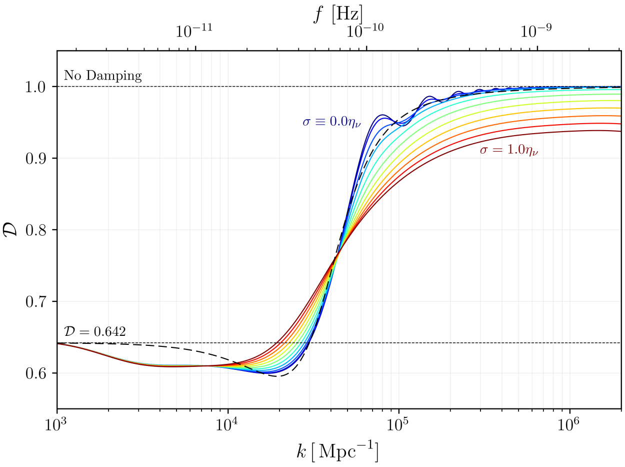


Figure 6. Illustration for the damping envelope at large k. At $k\gtrsim3\times10^5\,{\rm Mpc}^{-1}$, the envelope has essentially returned to the undamped solution, while by $k\simeq10^3\,{\rm Mpc}^{-1}$ the waves reach the expected $\mathcal{D}=0.642$. A parameter σ modulates the *sharpness* with which the neutrino decoupling occurs. Each increment in colour corresponds to increasing σ by 0.1. The dashed line shows the simple approximation given in Eq. (21).

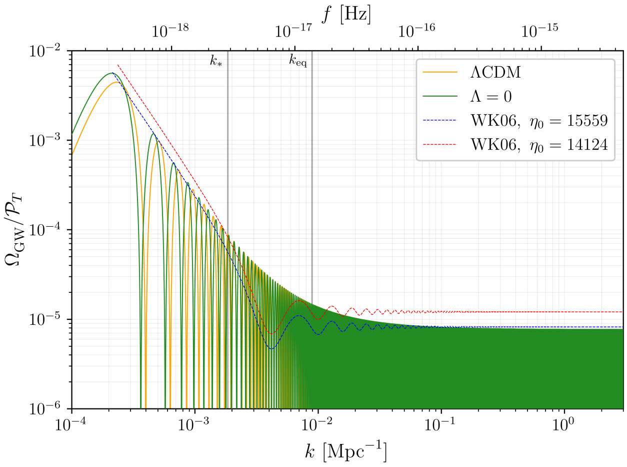


Figure 7. A graph showing $\Omega_{\rm GW}/\mathcal{P}_T$ across k as seen at η_0 in two different Universes: a standard ΛCDM and one without a cosmological constant. Differences are only present for the smallest k values, and correspond to phase shifts rather than a difference in fundamental spectral shape.

5.1.1 *Cosmology dependence of the damping coefficient*

The total amplitude of the damping carries Cosmological dependence in the form of $f_{\nu,0}=\Omega_\nu/(\Omega_\gamma+\Omega_\nu)$, as defined in Eq. (17a), which in turn will depend on $N_{\rm eff}$. Using the iterative procedure for the damping contributions (see Sect. 4) we find a ΛCDM value of $\mathcal{D}=0.642$, differing slightly from Weinberg (2004), where it was concluded that $\mathcal{D}=0.644$ by using $N_{\rm eff}=3$, implying $f_{\nu,0}=0.40523$. By running the solution for values $0\leqslant f_{\nu,0}\leqslant1$ we find the fit

$$\mathcal{D}=1-0.45397\,\zeta+0.11375\,\zeta^2-0.01904\,\zeta^3+0.00168\,\zeta^4 \qquad (22)$$

where $\zeta=f_{\nu,0}/0.40890$. This choice of pivot value is derived from the theoretically expected $N_{\rm eff}=3.046$, since the current measured value is poorly constrained to $N_{\rm eff}=2.99\pm0.17$ (Planck Collaboration et al. 2018b). Inserting $f_{\nu,0}=0.40523$, we obtain $\mathcal{D}=0.6449$, which more closely matches the value of Weinberg (2004), and matches the $\mathcal{D}=0.645$ of DR04, based on the same $f_{\nu,0}$, thus confirming the equivalence of our treatments.

5.2 Late time acceleration

The inclusion of a DE component in the Universe's expansion does not allow for simple analytic expressions like those given in Eq. (5). Despite this, we can argue that the effects of a cosmological constant will be small, since this component only becomes dominant at very late times. The scale factor at which Ω_Λ matches the contribution from $\Omega_{\rm m}$ is given by

$$a_\Lambda = \left(\frac{\Omega_{\rm m}}{\Omega_\Lambda}\right)^{1/3}, \tag{23}$$

which takes a value of $a_\Lambda \simeq 3/4$ with current best-fit parameters (Planck Collaboration et al. 2018b), a value close to today's scale factor $a_0 = 1$ (e.g. see vertical lines in Fig. 2). Recalling that the effects of cosmological expansion are imprinted on the GWB at the time of horizon crossing, this means that only large scales which crossed horizon recently can be impacted in spectral shape. This can be verified with the numerical solution – where arbitrary expansion histories are easily included – as can be seen in Fig. 7. We see that the spectral shape is unchanged, with only specific local maxima showing a shift of position, and most of the spectrum simply receiving a phase shift. Sensitivity to the shifted peaks would require sensitivity to wavelengths spanning large fractions of the observable Universe, which even if feasible would be heavily limited by cosmic variance. Similarly the phase shift is invisible to probes which typically average the spectrum over one or many cycles. This suggests therefore that late acceleration can be neglected for practical purposes.

Although the shape of the spectrum does not change significantly, one important effect is in reducing the value of η_0. In words, if one includes late time accelerated expansion then the waves have less time to evolve before the scale factor reaches today's value of $a_0 = 1$. This means that when using the approximations given in Eq. (9) one should use the *wrong* value $\eta_0 \approx 15560\,{\rm Mpc}$ for more accurate results in a full ΛCDM Universe. Using the *correct* value of $\eta_0 \approx 14120\,{\rm Mpc}$ gives a correspondingly younger spectrum, and hence overall larger amplitude. This is shown in Fig. 7, where we depict the expected limits of the analytic solution by plotting an interpolated line of the local peaks.

Another potential cancellation of errors arises here. By inspecting the second line in Eq. (15) we see that, with H_0 held constant, the fundamental dependence of the energy spectrum is $\propto \eta_*^2/\eta_0^4$. A cancellation of errors, which in fact gives the correct result to within $\simeq 15\%$, is to use the lower η_* from neglecting neutrinos, with the lower η_0 from a younger late-accelerated universe. This again makes diagnosing discrepancies in the literature difficult, especially if there is ambiguity between $\Omega_{\rm rel}$ and Ω_γ.

5.3 Relativistic degrees of freedom

Accounting for the changing relativistic degrees of freedom involves modifying the evolution of the scale factor (e.g., Watanabe & Komatsu 2006, see):

$$a' = a^2 H_0 \sqrt{\frac{g_{*\rho}}{g_{*\rho 0}}\left(\frac{g_{*s0}}{g_{*s}}\right)^{4/3}\Omega_{\rm rel}a^{-4} + \Omega_{\rm m}a^{-3} + \Omega_\Lambda}. \tag{24}$$

The functions for $g_{*\rho}$ and g_{*s} are available in pretabulated or functional forms, together with much more detailed discussion of the physics at play, in Saikawa & Shirai (2018).

This change in the energy budget can be interpreted as a departure from the expected $\rho \propto a^{-4}$ behaviour of relativistic fluids, but only in specific temperature ranges where there is some change in the thermodynamics of the plasma, e.g., during phase transitions. These small changes in the expansion rate will be imprinted on the GWB, as illustrated in Fig. 8, where we again numerically solved the transfer functions with modified expansion rates. It is noteworthy that this change in the energy budget of relativistic particles changes the exact relation between conformal time η, scale factor a, and temperature T.

Fortunately, the lowest frequencies impacted by changes in the relativistic energy budget are $k \simeq 10^3\,{\rm Mpc}^{-1}$, which is deep into the region of tensor modes which entered horizon during RD. This means the limiting case of $\langle\Omega_{\rm GW}\rangle = (1/2)\,\Omega_{\rm rel}/12$ in the absence of any other physical effects can be safely extrapolated (see discussion at the end of Sect. 4.2). In particular an extra factor which approximately accounts for the changes in the relativistic degrees of freedom is given by (Saikawa & Shirai 2018)

$$\langle\Omega_{\rm GW}^{g_*}\rangle \approx \langle\Omega_{\rm GW}\rangle\left(\frac{g_\rho}{g_{\rho 0}}\right)\left(\frac{g_{s0}}{g_s}\right)^{4/3}. \tag{25}$$

Using the ΛCDM envelopes given in Table 1, noting that the ν damped envelope requires the modification in Eq. (21), matches the full solution to within $\simeq 5\%$ as measured with respect to our full numerical solution. This means that the relatively simple fits in this paper together with the pretabulated g_* functions (see specifically Appendix A in Saikawa & Shirai (2018)) the entire ΛCDM GWB can be replicated to high precision across a very large range of scales.

The g_* effects discussed here were again neglected in previous calculations of the spectral distortion window functions[11]. If the effects were included they would remove a similar percentage of sensitivity as the damping effects, but over a different range of scales. This combined with the over-extension of damping effects will lead to almost unchanged results in Kite et al. (2020), but should be fully accounted for in future distortion studies.

6 DISCUSSION AND CONCLUSION

The GWB offers an exciting new window to the physics of the early Universe, and a diverse set of probes will soon begin the search for this new signal. In this paper we give simple yet precise large-scale functional forms for the GWB energy density, which can aid in estimating the efficacy of some observations. The scales considered here are especially helpful in comparing CMB *B*-modes and CMB spectral distortion measurements (as required in Kite et al. 2020). More generally, however, any comparison between early- or late-universe probes requires understanding of the η_* scale for MD–RD transition.

In this paper we endeavoured to firstly elucidate the physics at play in the GWB in a pedagogical way, secondly to provide tools for simple calculation of the large-scale energy spectrum, and finally discuss the main features in the GWB with special attention on consequences for spectral distortion calculations.

In Sect. 2 we qualitatively review the physical phenomena affecting the large-scale spectrum. The most important of these is the transition between a universe dominated by relativistic particle species and matter. We explicitly discuss the importance of including neutrinos, despite the common misnomer of *radiation domination*. A second important effect we discuss is that of neutrino damping via anisotropic stress in the medium. In Sect. 3 we give solutions to the transfer function valid in RD and MD respectively [see

---

[11] Note however that some of the models discussed in Kite et al. (2020) implicitly included the g_* effects in their energy spectra, and were thus indirectly included in the distortion calculation. In future these should be included in the window function itself for completeness and higher accuracy.

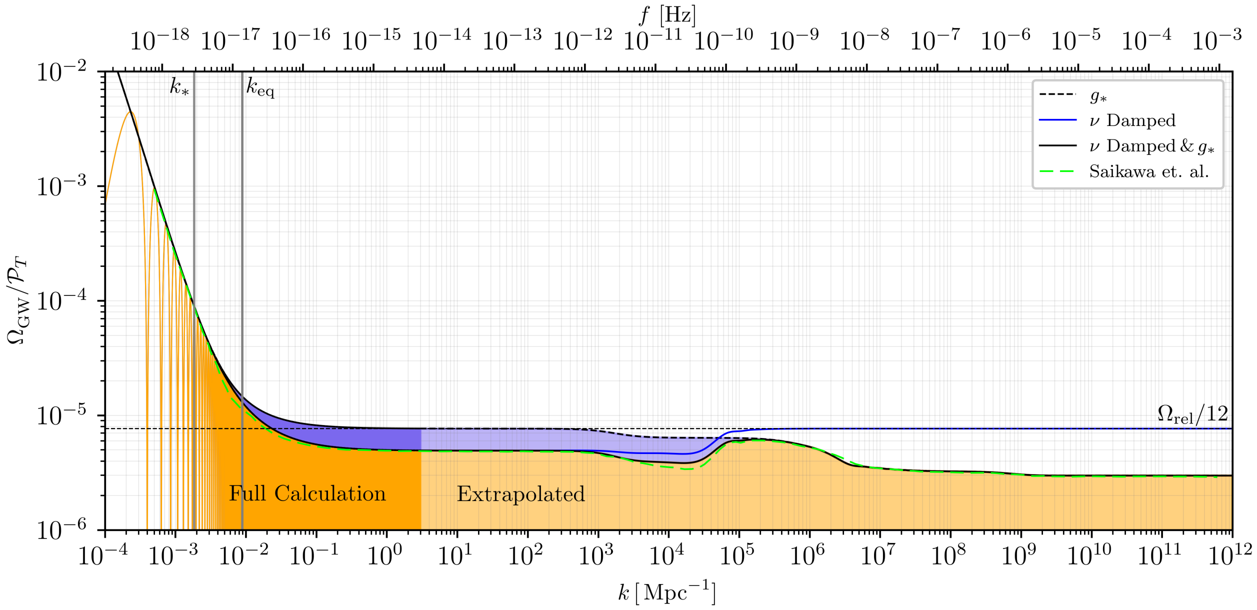


Figure 8. A graph showing the GWB for an extremely wide range of scales, extending to the highest frequencies interferometry missions aim to measure, and also the highest energy scales that known physics allow us to model. We show the onset of the effects of damping and the effects of changing relativistic degrees of freedom. For comparison we show digitalised data from Saikawa & Shirai (2018), which matches well to the findings here.

Eqs. (9),(10)], and explain how these are matched at the transition assuming this to be instantaneous [Eq. (12)]. These solutions give expected limits for the energy density of the GWB [Eqs. (15),(16)]. In Sect. 4 we explain how the damping can be treated through an iterative numerical method. This method is found to give results matching those of WK06 and DR04 in the appropriate limits (see Fig. 2). Energy spectra as seen today are shown for various limiting cases: with and without damping, with and without DE, with and without neutrinos (see Figs. 5,7). To replicate these spectra with ease, we provide coefficients in Table 1 for use with Eq. (19), which is valid to scales of $k \simeq 10^3\,\rm Mpc^{-1}$. Sect. 5 finalises the analysis with discussion of various GWB features: cosmological dependence of neutrino damping, effects of late time accelerated expansion, and changes in the number of relativistic degrees of freedom.

The effects included in this analysis were purely standard model Physics. More generally one would apply the techniques discussed here to verify the avenues of discovery for new Physics hidden in the GWB. We note that the numerical method utilized in this work can be straightforwardly generalized to non-standard thermal histories that transition from RD to MD and back one or more times, as is the case in a variety of scenarios of beyond the standard model physics (Acharya et al. 2008, 2019; Arbey et al. 2021). It can also be extended to include the presence of other light, weakly interacting particles, such as axions or axion-like particles in the early universe (Marsh 2016). We defer the implementation of this to a future study.

Two of the features explored in this paper reveal inaccuracies in previous calculations of tensor window functions, $W_\mu(k)$, used to calculate SD amplitudes arising from primordial tensor power spectra (e.g., Chluba et al. 2015). Previously damping has been included, but extending to arbitrarily high k. The damping ceases to affect the spectrum beyond $k \simeq 10^5\,\rm Mpc^{-1}$ (see Fig. 6). However, the effects of the relativistic degrees of freedom were also not included explicitly within W_μ, and would lead to consecutive over- and underestimations on scales $k \gtrsim 10^3\,\rm Mpc^{-1}$. Together all these changes add to only small percent changes, rendering the conclusions of Kite et al. (2020) still valid. The calculation of new and more precise window functions remains as future work, where it would be appropriate to include the tensor perturbations within a full Boltzmann code, and accurately model the GW-photon interaction, and thus fully capturing the smooth decoupling of the photons, even in the post-recombination era.

DATA AVAILABILITY

Data in all figures available at `https://doi.org/10.5281/zenodo.5141789`. A full release of the `CosmoTherm` code is planned for the near future, including the GW module used here.

ACKNOWLEDGMENTS

We would like to thank Eiichiro Komatsu and Yuki Watanabe for valuable comments on the draft. We also thank Paolo Campeti for helpful discussions on the role of η_0 on the approximate spectrum. Finally we thank Chiara Caprini and Daniel Figueroa for their encouraging words on the draft.

This work was supported by the ERC Consolidator Grant *CMBSPEC* (No. 725456) as part of the European Union's Horizon 2020 research and innovation program. TK was further supported by STFC grant ST/T506291/1. JC was also supported by the Royal Society as a Royal Society URF at the University of Manchester, UK.

REFERENCES

Abbott B. P., et al., 2020a, ApJ, 892, L3
Abbott R., et al., 2020b, ApJ, 896, L44
Acharya B. S., Dhuria M., Ghosh D., Maharana A., Muia F., 2019, JCAP, 11, 035

Acharya B. S., Kumar P., Bobkov K., Kane G., Shao J., Watson S., 2008, JHEP, 06, 064

Ade P., et al., 2018, Phys. Rev. Lett., 121, 221301

Aghanim N., et al., 2020, Astron. Astrophys., 641, A6

Ahmad Q. R. et al., 2002, Physical Review Letters, 89, 011301

Ahmad Q. R. et al., 2001, Physical Review Letters, 87, 071301

Aker M., et al., 2021

Alam M. F. et al., 2020, arXiv e-prints, arXiv:2005.06490

Arbey A., Auffinger J., Silk J., 2021, PoS, ICHEP2020, 585

Ballesteros G., Rey J., Rompineve F., 2020, JCAP, 06, 014

Baym G., Patil S. P., Pethick C. J., 2017, Phys. Rev. D, 96, 084033

Bhattacharya S., Das A., Dutta K., 2021

Boyle L. A., Steinhardt P. J., 2008, Phys. Rev. D, 77, 063504

Campeti P., Komatsu E., Poletti D., Baccigalupi C., 2021, J. Cosmology Astropart. Phys., 2021, 012

Caprini C., Figueroa D. G., 2018, Class. Quant. Grav., 35, 163001

Chluba J., 2014, MNRAS, 440, 2544

Chluba J., Dai L., Grin D., Amin M. A., Kamionkowski M., 2015, MNRAS, 446, 2871

Chluba J., Sunyaev R., 2012, Mon.Not.Roy.Astron.Soc., 419, 1294

de Salas P. F., Pastor S., 2016, J. Cosmology Astropart. Phys., 2016, 051

Dicus D. A., Repko W. W., 2005, Phys. Rev. D, 72, 088302

Flauger R., Weinberg S., 2018, Phys. Rev. D, 97, 123506

Fukuda Y. et al., 1998, Physical Review Letters, 81, 1562

Green A. M., Kavanagh B. J., 2021, J. Phys. G, 48, 4

Jeong D., Pradler J., Chluba J., Kamionkowski M., 2014, Physical Review Letters, 113, 061301

Kite T., Ravenni A., Patil S. P., Chluba J., 2020, arXiv e-prints, arXiv:2010.00040

Lesgourgues J., Pastor S., 2006, Phys. Rept., 429, 307

LIGO Scientific Collaboration, Virgo Collaboration, 2016, Physical Review Letters, 116, 061102

Ma C.-P., Bertschinger E., 1995, ApJ, 455, 7

Mangano G., Miele G., Pastor S., Pinto T., Pisanti O., Serpico P. D., 2005, Nuclear Physics B, 729, 221

Marsh D. J. E., 2016, Phys. Rept., 643, 1

Mirón-Granese N., 2020

Perera B., et al., 2019, Mon. Not. Roy. Astron. Soc., 490, 4666

Planck Collaboration et al., 2018a, arXiv e-prints, arXiv:1807.06205

Planck Collaboration et al., 2018b, ArXiv:1807.06209

Saikawa K., Shirai S., 2018, J. Cosmology Astropart. Phys., 2018, 035

Watanabe Y., Komatsu E., 2006, Phys. Rev. D, 73, 123515

Weinberg S., 2004, Phys. Rev. D, 69, 023503

Weinberg S., 2008, Cosmology. Oxford University Press

Zarei M., Bartolo N., Bertacca D., Matarrese S., Ricciardone A., 2021

6

Paper III: Spectro-spatial evolution of the CMB I: discretisation of the thermalisation Green's function

This paper is the first in a series of three papers (see chapter 7 and chapter 8) which together aim to correctly model the frequency domain for the photon phase space distribution (see chapter 2) within the Boltzmann hierarchy (see chapter 3).

The first step in this new formalism is to build a basis of spectral shapes which taken together can discretise the frequency space such that thermalisation can be appropriately modelled. To this end we define the boosted spectral shapes formed from applying $x\partial_x$ to the y-distortion spectral shape $\mathcal{Y}_{\rm SZ}(x)$, understanding that this boosting operator is an important aspect of early Universe perturbation theory (see chapter 3).

We show that in this basis we can model both energy exchanging photon interactions through the Kompaneets operator and the emission or absorption of photons (again see chapter 2). The theoretical arguments are supplemented with a comparison to the full solutions offered by binning the frequency space and running a *brute force* calculation (Chluba, 2013, 2015), which shows accurate results. Importantly this means replicating the full solutions with $100 - 1000\times$ less computational complexity, and also having physically representative and meaningful basis functions which directly map to the anisotropic calculation explored in the following two papers.

Prepared for submission to JCAP

Spectro-spatial evolution of the CMB I: discretisation of the thermalisation Green's function

Jens Chluba[a] , **Thomas Kite**[a] and **Andrea Ravenni**[b,c]

[a] Jodrell Bank Centre for Astrophysics, School of Physics and Astronomy, The University of Manchester, Oxford Road, Manchester, M13 9PL, U.K.
[b] Dipartimento di Fisica e Astronomia "Galileo Galilei", Università degli Studi di Padova, via Marzolo 8, I-35131, Padova, Italy.
[c] INFN, Sezione di Padova, via Marzolo 8, I-35131, Padova, Italy.

E-mail: Jens.Chluba@Manchester.ac.uk, Thomas.Kite@Manchester.ac.uk, Andrea.Ravenni@unipd.it

Abstract. Spectral distortions of the cosmic microwave background (CMB) have been recognized as an important future probe of the early Universe. Existing theoretical studies primarily focused on describing the evolution and creation of *average distortions*, ignoring spatial perturbations in the plasma. One of the main reasons for this choice is that a treatment of the spectro-spatial evolution of the photon field deep into the primordial Universe requires solving a radiative transfer problem for the distortion signals, which in full detail is computationally challenging. Here we provide the first crucial step towards tackling this problem by formulating a new spectral discretisation of the underlying average thermalisation Green's function. Our approach allows us to convert the high-dimensional partial differential equation system ($\simeq 10^3 - 10^4$ equations) into and set of ordinary differential equations of much lower dimension ($\simeq 10$ equations). We demonstrate the precision of the approach and highlight how it may be further improved in the future. We also clarify the link of the observable spectral distortion parameters (e.g., μ and y) to the computational spectral basis that we use in our frequency discretisation. This reveals how several basis-dependent ambiguities can be interpreted in future CMB analysis. Even if not exact, the new Green's function discretisation can be used to formulate a generalised photon Boltzmann-hierarchy, which can then be solved with methods that are familiar from theoretical studies of the CMB temperature and polarisation anisotropies. We will carry this program out in a series of companion papers, thereby opening the path to full spectro-spatial exploration of the CMB with future CMB imagers and spectrometers.

Contents

1 Introduction **2**

2 Approximate ODE representation of the thermalisation Green's function **3**
- 2.1 Brief recap of the thermalisation Green's function 4
- 2.2 Basic idea and lowest order solution of the thermalisation problem 5
- 2.3 Preliminaries 6
- 2.4 Spectral basis and approximate representation of the Kompaneets operator 7
- 2.5 Extension of the basis 9
- 2.6 Generalization of the ODE system 9
- 2.7 Adding the effect of photon production and heating 12
- 2.8 Solutions for the Green's function after single injection 13

3 Defining spectral distortion observables **15**
- 3.1 Efficient change of the basis 17
- 3.2 Performance and convergence 18
- 3.3 Caveats of the observation basis and alternative description 19

4 Discussion and conclusions **22**

A Useful operator properties **28**

B Alternative derivation of the ODE system **30**

1 Introduction

Spectral distortions (SDs) of the cosmic microwave background (CMB) have now been recognized as an important future probe of early-universe and particle physics. In particular the ability of CMB SDs to constrain the primordial power spectrum at small scales [1–4] provides important motivation to push the observational frontier with the next generation CMB experiments [5, 6]. However, much of the recent theoretical work [e.g., 7–11] and experimental spectrometer concept studies [12–17] focused primarily on the science of *average distortion signals*. It is known however that an average distortion signal can form anisotropies through interaction with other perturbed quantities [4]. While average distortion physics is often concerned with earlier times and thus sourced by larger k-modes, the anisotropies formed from this average distortion would be prevalent on similar scales to usual CMB anisotropies [explored further in 18, 19]. From the theoretical point of view it is clear that distortion anisotropies should be smaller and thus harder to detect, however, one of the main reasons for preferring average distortion science is that the computations of distorted CMB anisotropies of primordial origin are difficult and currently beyond the possibilities of existing Boltzmann codes.

To illustrate this statement we highlight that numerically solving the full thermalisation problem for SDs created on average by energy release now takes of order $\simeq 30$ seconds on a standard laptop using `CosmoTherm` [7, 20]. While this is already highly optimised, it will be difficult to extend this method to SD anisotropies, where in analogy to the standard CMB temperature fluctuations [21, 22] one would have to solve the thermalisation problem for multiple k-modes. For each k-mode, a multipole hierarchy would furthermore be required, overall boosting the computation by a factor $\simeq 10^3$. In addition, one would have to consider how to convert the final (frequency-dependent) signal transfer functions into CMB observables, which further increases the complexity of the problem over the standard CMB anisotropy computation, likely yielding single computations that would take $\mathcal{O}(10^5 - 10^6)$ seconds. While not necessarily prohibitively expensive with modern computational resources, this brute force approach would be overly-complicated for exploratory calculations and not scalable in parameter forecasts and searches for new physics.

How could one make the problem more tractable? The most common approach is to simplify the problem by considering limiting cases. In particular, scenarios in which the evolution of distortions and primordial perturbations as well as thermalisation physics can be mostly separated come to mind. This brings us to the well-known Sunyaev-Zeldovich (SZ) effect [23, 24], which is created by anisotropic heating effects in the late Universe, sourcing y-type distortion anisotropies that peak at several arcminute angular scales. This signal is highly non-Gaussian and requires an understanding of the non-linear large-scale structure evolution, but then analytically translates the statistical properties of the dark matter distribution into the y-field [25–29]. The SZ effect is therefore an important probe for cosmology and cluster physics [30, 31].

Another example is the sourcing of y-distortion anisotropies by the mixing of blackbodies in the perturbed universe [1, 2]. This second order effect leads to a fluctuating y-distortion sky [32, 33] in addition to an average distortion [4] when perturbations dissipate by free-streaming and Thomson scattering effects. For the fluctuating part, no spectral evolution has to be considered at the late stages (redshift $z \lesssim 10^4$), just like for the SZ effect – a linear perturbation description of the problem is furthermore possible, yielding y-parameter transfer functions that are excited by first order temperature perturbations [33, 34]. If the amplitude of the small-scale curvature perturbations is modulated by large-scale modes this can furthermore lead to correlated $\mu \times T$ and $y \times T$ fluctuations [35–42], which can be directly constrained using CMB imagers [see 43–45, for most recent forecasts and constraints]. Note that at the largest angular scales, the corresponding transfer problem was simplified by neglecting details of the distortion evolution in the perturbed Universe [35, 40, 46].

There are, however, a number of aspect to the thermalisation problem that have not been captured by any of these calculations. As explained in [4], if an average distortion is present during the pre-recombination era, the standard density perturbations at first order will source distortion anisotropies. Assuming the average SD is $\Delta n_\nu^{(0)}$ in terms of the photon occupation number, the SD anisotropies will have a spectrum that follows $\Delta n_\nu^{(1)} \propto -\nu\partial_\nu \Delta n_\nu^{(0)}$ [4]. Even without any spectral evolution, the standard Doppler terms and potential perturbations therefore source distortion anisotropies, which have not been evaluated for a more general average spectrum [see 33, for a treatment of *occupation number* temperature and y-distortion]. Assuming that the average distortion saturates the limits imposed by *COBE/FIRAS* [47, 48], one can expect distortion anisotropies at the level of $\simeq 10^{-8}$–10^{-7} of the average CMB. This can exceed the signals expected from the aforementioned non-Gaussian signals and can also be directly constrained with existing and future CMB imaging data. In addition, the thermalisation efficiency should vary from patch to patch in the perturbed Universe. The required terms in the photon Boltzmann equation were already discussed in [4]; however, only recently has the effect been estimated using a *separate universe* approach [49]. In particular for modes that cross the horizon at or after the recombination process completes this effect should be noticeable in the transfer function solutions, but has not been computed using a full Boltzmann treatment.

To fully capitalise on the potential of spectral distortion anisotropy studies, as a first step we need to formulate a generalized photon Boltzmann equation that goes beyond the standard temperature and polarisation anisotropies. The biggest bottleneck is due to the discretisation of the spectral evolution, which currently is done with $\simeq 10^3$–10^4 bins in frequency, as explained above. In this work, we obtain a new discretisation for the average frequency evolution that reduces the computational burden by a factor of $\simeq 10^3$ (Sect. 2). This allows us to model the thermalisation from $y \rightarrow \mu \rightarrow T$ with a small number ($\simeq 10$) of new spectral parameters, that can represent the exact calculation from `CosmoTherm` to high precision. In contrast to other approximations, the solution is no longer limited to the three standard spectral shapes but allows one to capture the dominant contributions from the residual distortion [e.g., 50]. We also explain how the computational distortion parametrisation can be mapped back onto the leading residual distortion spectra, which present the main spectral shapes that may be testable in future applications (Sect. 3). we close with a brief discussion of the next steps and remaining limitations in Sect. 4.

This paper is the first in a series of works that study the effect of spectro-spatial evolution of the CMB. In paper II [18], we will formulate the generalised Boltzmann equation for distortions cause by heating processes, strongly drawing on the results of this paper. In paper III [19], we will present a detailed discussion of the distortion transfer functions and power spectra, highlighting the importance of various physical effects and providing Fisher forecasts. We also plan subsequent papers that discuss how the dissipation of acoustic modes in the presence of primordial non-Gaussianty causes spectral distortion anisotropies, and which constraints on various scenarios can be expected. Overall we hope this will provide further motivation to study SDs in the future.

2 Approximate ODE representation of the thermalisation Green's function

In this section, we establish a novel way of modeling the spectral evolution of the average photon field under repeated Compton scattering and thermal photon emission processes. In terms of perturbation theory, this is akin to focusing on the background quantities only, which leads back to the Green's function approach for the thermalisation problem, as will be developed here. Anisotropic distortion evolution from heating processes will be considered in papers II and III of the series.

2.1 Brief recap of the thermalisation Green's function

The efficiency of photon production and Comptonisation in the primordial plasma dictate various eras with characteristic SD shapes which are defined below. At sufficiently early times, in the temperature or T-era ($2 \times 10^6 \lesssim z$), thermalisation processes are very efficient and any excess energy is rapidly converted into a temperature shift, $G(x)$. Here $x = h\nu/kT_z$ where $T_z = T_0(1+z)$ is the background reference temperature, which is chosen to match today's CMB temperature $T_0 = 2.7255\,\mathrm{K}$ [51].[1] The subsequent μ-era ($5 \times 10^4 \lesssim z \lesssim 2 \times 10^6$) is characterised by a lack of photon production, leading to a chemical potential distortion, $M(x)$. Finally, the y-era ($z \lesssim 5 \times 10^4$) renders photon energy redistribution inefficient, leading to a distortion, $Y(x)$, related to the well-known SZ effect, albeit in this case of primordial origin.

The different characteristic spectra introduced above have the forms [e.g., 10, 52]

$$G(x) = \frac{x\,\mathrm{e}^x}{(\mathrm{e}^x - 1)^2}, \qquad Y(x) = G(x)\left[x\frac{\mathrm{e}^x + 1}{\mathrm{e}^x - 1} - 4\right], \qquad M(x) = G(x)\left[\frac{1}{\beta_M} - \frac{1}{x}\right], \tag{2.1}$$

with $\beta_M = 3\zeta(3)/\zeta(2) \approx 2.1923$. In Sect. 2.4 we will describe how these can be obtained by boosts of the average blackbody spectrum. Central properties of these spectra are summarised by their dimensionless photon number density $N_f = \int x^2 f(x)\,\mathrm{d}x$ and energy density $E_f = \int x^3 f(x)\,\mathrm{d}x$. The corresponding integrals can be carried out analytically in terms of Riemann ζ-functions. Also using the blackbody occupation number, $n_{\rm bb}(x) = 1/(\mathrm{e}^x - 1)$, we then find:

$$N_{n_{\rm bb}} = 2\zeta(3) \approx 2.40411, \qquad E_{n_{\rm bb}} = \frac{\pi^4}{15} \approx 6.49394 \tag{2.2a}$$

$$N_G = 6\zeta(3) = 3N_{n_{\rm bb}} \approx 7.21234, \qquad E_G = E_Y = \frac{4\pi^4}{15} = 4E_{n_{\rm bb}} \approx 25.9758 \tag{2.2b}$$

$$N_Y = N_M = 0, \qquad E_M = \frac{2\pi^6}{135\zeta(3)} - 6\zeta(3) \approx \frac{E_{n_{\rm bb}}}{1.40066} \approx 4.63635. \tag{2.2c}$$

The absence of overall photon number for y and μ type distortions is by construction (and easily achieved by subtracting G from alternative definitions). This convention has already been commonplace in the literature, but will become a fundamental simplifying fact in the novel treatment introduced below.

While the heuristic decomposition into three distinct eras introduced above conveys the correct physics to relatively high precision, it is much more convenient to have a robust framework in which the results can be expanded and built upon. In [53] it was shown that the thermalisation problem can be expressed as a Green's function problem in the limit of small energy injection:

$$s = \alpha_f \int_0^\infty \mathcal{J}_s(z)\,\frac{\mathrm{d}Q}{\mathrm{d}z}\,\mathrm{d}z, \tag{2.3}$$

where $s \in \{\Theta \equiv \Delta T/T_z, \mu, y\}$ gives the signal amplitude of the corresponding SD ($f \in \{G, Y, M\}$), $\mathcal{J}_s$ is a dimensionless energy branching ratio, and $\alpha_f \equiv E_{n_{\rm bb}}/E_f \in \{1/4, 1/4, 1.40066\}$ is an energetic conversion factor from a blackbody spectrum to the SD amplitude [easily read off from Eq. (2.2)]. The energy release is determined by the comoving relative energy injection rate, $\frac{\mathrm{d}Q}{\mathrm{d}z} = \frac{1}{\rho_\gamma}\frac{\mathrm{d}Q_{\rm c}}{\mathrm{d}z}$, where $\mathrm{d}Q_{\rm c}/\mathrm{d}z$ directly follows from the photon collision term.

For clarity we note that the *three era* picture of the early Universe would correspond to simple top-hat functions for $\mathcal{J}_s$ [i.e., see 'Method A' in 54]. Other approximations for the energy branching ratios of varying accuracy exist [54], including the addition of intermediate spectral shapes known as *residual distortions* [50] and perturbative SD approximations for moderate scattering y-parameter [55]. A byproduct of this work is the ability to generate accurate Green's functions in a generalized spectral basis to a precision comparable with full numerical treatments (see Sect. 2.2).

---

[1] This avoids having to deal with redshifting terms.

While the spectral shapes in Eq. (2.1) are physically motivated – each characteristic of a limiting case for each phase in the early Universe – they are insufficient to model the general evolution of the spectrum caused by heating processes. In the following sections we introduce a method for extending this set of spectral functions and explain how this new spectral basis eventually allows for full spectro-spatial solutions of primordial perturbations in the photon field.

2.2 Basic idea and lowest order solution of the thermalisation problem

As previously mentioned, in the μ-era all injected energy rapidly converts into the μ-distortion [1, 56–58]. The net μ-parameter is given by the evolution equation $\frac{\partial \mu}{\partial t} \approx \gamma_\rho \dot{Q}$, where $\gamma_\rho \equiv \alpha_M \approx 1.4007$ and $\dot{Q} = \mathrm{d}Q/\mathrm{d}t$. For a given $\dot{Q}$, this equation can be solved with initial $\mu = 0$.

Physically, the energy injection first leads to an increase in the distortion y-parameter by $\dot{y} \approx \frac{1}{4}\dot{Q}$, which then quickly converts into μ. If we insert this intermediate step, we may instead write

$$\frac{\partial y}{\partial t} \approx \frac{1}{4}\dot{Q} - 4\dot{\tau}\theta_z\, y \qquad \text{and} \qquad \frac{\partial \mu}{\partial t} \approx \gamma_\rho\, 4\dot{\tau}\theta_z\,(4y) \approx 22.411\,\dot{\tau}\theta_z\, y. \tag{2.4}$$

Here, $4y$ is the relative momentary energy density within the y-distortion part, $\theta_z = k_\mathrm{B} T_z/m_\mathrm{e}c^2$ is the dimensionless temperature,[2] and $\dot{\tau} = \mathrm{d}\tau/\,\mathrm{d}t = N_\mathrm{e}\,\sigma_\mathrm{T} c$ denotes the differential Thomson optical depth, all with the common choice of constants. The average energy exchange rate is $\langle \Delta\nu/\nu \rangle \simeq 4\theta_z$ per scattering [59, 60], which determines how quickly energy flows from y to μ. This identifies $\dot{\tau}\theta_z$ as a fundamental timescale in the thermalisation problem which contrasts with the timescale of Thomson scattering $\dot{\tau}$ – a fact which will become important for the generalised Boltzmann hierarchy (paper II). As we see in Fig. 1, the solution of this simple system roughly captures the transition between the μ and y-eras, yielding the y-distortion visibility $\mathcal{J}_y \approx \mathrm{e}^{-4y_z}$, where we introduce the scattering y-parameter $y_z = \int \dot{\tau}\theta_z\,\mathrm{d}t$, and, from energy conservation, the μ-visibility $\mathcal{J}_\mu \approx 1 - \mathcal{J}_y$.

Following [61], the reduction of the chemical potential by the Bremsstrahlung (BR) and double Compton (DC) processes is approximately given by $\dot{\mu}|_\mathrm{em/abs} \approx -\gamma_N\, x_\mathrm{c}\, \dot{\tau}\theta_z\, \mu$ with $\gamma_N \approx 0.7769$. Here, x_c is the critical frequency between DC/BR emission and Compton scattering, with approximations as a function of redshift given in Sect. 3.3.1 of [11]. Every absorption event then removes $\Delta\dot{\rho}_\gamma/\rho_\gamma \approx \alpha_M^{-1}\,\dot{\mu}|_\mathrm{em/abs}$ of energy from the μ-distortion, which is immediately added back to the average temperature, causing a relative temperature shift $\Theta = \Delta T/T_z$. Assuming energy conservation, we therefore have the corresponding temperature source term $\dot{\Theta}|_\mathrm{em/abs} = -\frac{1}{4\gamma_\rho}\,\dot{\mu}|_\mathrm{em/abs} \approx \frac{\gamma_N}{4\gamma_\rho}\, x_\mathrm{c}\, \dot{\tau}\theta_z\, \mu$. Overall this means one has to solve the extended system

$$\frac{\partial \Theta}{\partial t} \approx \gamma_T\, x_\mathrm{c}\, \dot{\tau}\theta_z\, \mu, \qquad \frac{\partial y}{\partial t} \approx \frac{1}{4}\dot{Q} - 4\dot{\tau}\theta_z\, y \qquad \text{and} \qquad \frac{\partial \mu}{\partial t} \approx \gamma_\rho\, 4\dot{\tau}\theta_z\,(4y) - \gamma_N\, x_\mathrm{c}\, \dot{\tau}\theta_z\, \mu, \tag{2.5}$$

with $\gamma_T = \frac{\gamma_N}{4\gamma_\rho} \approx 0.1387$, a task that can be easily carried out numerically. Assuming that the photon production process only becomes important when y is already negligible, it is easy to show that $\mathcal{J}_T \approx 1 - \mathcal{J}_\mathrm{bb}$ with distortion visibility $\mathcal{J}_\mathrm{bb} \approx \mathrm{e}^{-(z/z_\mu)^{2.5}}$ and $z_\mu = 1.98 \times 10^6$ [57, 58]. We then have

$$\mathcal{J}_y \approx \mathrm{e}^{-4y_z}, \qquad \mathcal{J}_\mu \approx (1 - \mathcal{J}_y)\,\mathcal{J}_\mathrm{bb}, \qquad \text{and} \quad \mathcal{J}_T \approx 1 - \mathcal{J}_\mathrm{bb}. \tag{2.6}$$

As can be seen from Fig. 1, these simple approximations already capture the main dependence of the distortion visibility on the injection redshift. The question of the next section is now whether we can improve on this description to also include terms relating to the residual distortion.

---

[2] We will use dimensionless temperatures, $\theta_X = kT_X/m_\mathrm{e}c^2$, frequently, with $T_X \in \{T_\mathrm{e}, T_z, T_\gamma\}$.

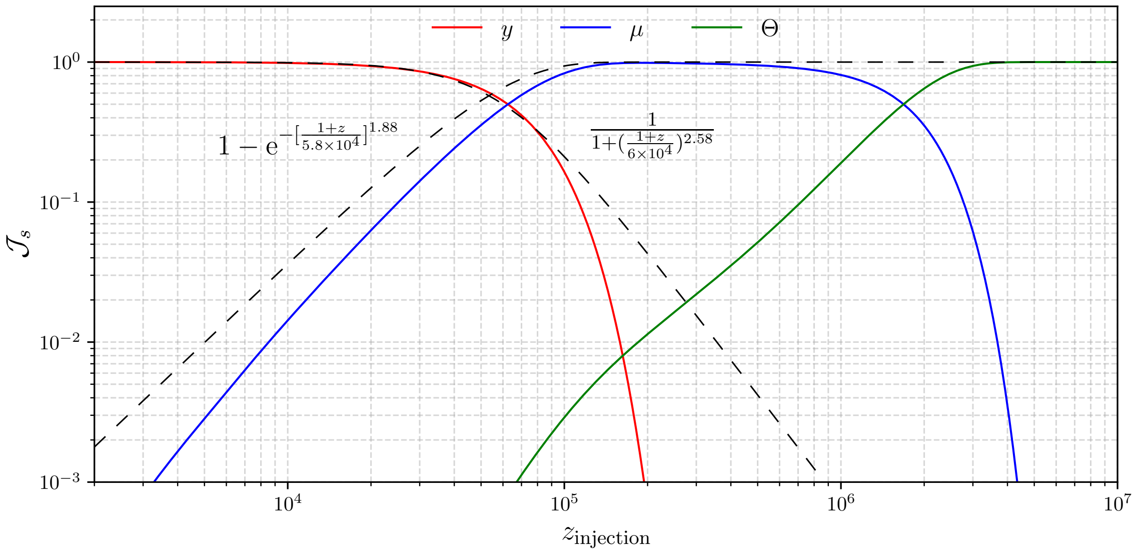


Figure 1: Fractions of energy in y, μ and Θ as seen today after injecting a normalised narrow Gaussian of energy at redshift $z_{\rm injection}$. This illustrates that the Green's function, as determined by Eq. (2.5), already broadly reproduces the least square fit results given in [53] based on the frequency-binned Green's function. The transition redshift, $z_{y\mu} \approx 5\times10^4$, to the μ-regime is obtained in both approximations. However, in comparison to the least square fits, $\mathcal{J}_y$ decays more rapidly towards high redshift.

2.3 Preliminaries

Neglecting photon production and heating terms, the relevant evolution equation for the distortion, $\Delta n = n - n_{\rm bb}$, from the blackbody $n_{\rm bb} = 1/({\rm e}^x - 1)$ in the expanding Universe can be cast into the compact form [7, 53]

$$\frac{\partial \Delta n(x)}{\partial y_z} \approx \Theta_{\rm e}\, Y(x) + \hat{\mathcal{K}}_x\, \Delta n(x) = \Theta_{\rm e}\, Y(x) + \hat{\mathcal{D}}_x\, \Delta n(x) + \hat{\mathcal{D}}^*_x\, A(x)\, \Delta n(x), \tag{2.7}$$

where $\Theta_{\rm e} = \frac{\Delta T_{\rm e}}{T_z}$ is the relative electron temperature difference and $\hat{\mathcal{K}}_x = \hat{\mathcal{D}}_x + \hat{\mathcal{D}}^*_x A$ is the Kompaneets operator, constructed from the diffusion and recoil operators, $\hat{\mathcal{D}}_x = x^{-2}\partial_x x^4 \partial_x$ and $\hat{\mathcal{D}}^*_x = x^{-2}\partial_x x^4$, with $A = 1 + 2n_{\rm bb} = ({\rm e}^x + 1)/({\rm e}^x - 1)$. The time variable is the scattering y-parameter as defined above. The problem has been linearised in the distortion, an approximation that will be good unless very large distortions are encountered [11, 20].[3]

For the electron temperature correction, $\Delta T_{\rm e} = T_{\rm e} - T_z$, we assume that Compton equilibrium is reached at all times.[4] In the absence of external heating, this means that $\int x^3 \partial_{y_z} \Delta n \,{\rm d}x \approx 0$, which implies $\Theta_{\rm e} \approx \Theta_{\rm eq}$ with [e.g., see 63]

$$\Theta_{\rm eq} \approx -\frac{\int x^3 \hat{\mathcal{K}}_x\, \Delta n\, {\rm d}x}{\int x^3 Y(x)\, {\rm d}x} = \frac{\int (x^4 \partial_x + x^4 A)\, \Delta n\, {\rm d}x}{4\, E_{n_{\rm bb}}} \equiv \frac{\int x^3 w_y\, \Delta n\, {\rm d}x}{4\, E_{n_{\rm bb}}} \tag{2.8}$$

---

[3] Note that the equations have not been linearized with respect to the reference blackbody, hence the presence of $A(x)$. This leads to blackbody-induced stimulated scattering [62], which is highly important for reaching equilibrium [e.g., 61].

[4] For the average evolution, this limit is valid on average until very late times corresponding to redshift $z \lesssim 200$.

and the y-weight factor $w_y = Y/G = xA(x) - 4 = x\frac{e^x+1}{e^x-1} - 4$. Since the integrals in Eq. (2.8) will appear multiple times, for convenience we introduce

$$\eta_f = \frac{\int x^3 w_y(x)\, f(x)\, \mathrm{d}x}{4\, E_{n_{\rm bb}}} \qquad \text{and} \qquad \epsilon_f = \frac{1}{\alpha_f} = \frac{\int x^3 f(x)\, \mathrm{d}x}{E_{n_{\rm bb}}}. \tag{2.9}$$

The exact integrals that are encountered in our computations can all be given in terms of the Riemann ζ-functions. For the basic spectral shapes we have $\eta_G = 1$, $\eta_Y \approx 5.3996$ and $\eta_M \approx 0.4561$, as well as $\epsilon_G = 4$, $\epsilon_Y = 4$ and $\epsilon_M = 1/1.4007$. For numerical applications we pre-compute all these integrals.

2.4 Spectral basis and approximate representation of the Kompaneets operator

The goal is to find an efficient spectral representation that captures the changes of the spectrum under repeated Compton scattering as described by Eq. (2.7). The simplest decomposition considers the three main spectral types appearing in the thermalisation problem introduced in Eq. (2.1). To build intuition, we discuss this case in some detail, but eventually find it is insufficient. The required refinements are presented right after.

It is instructive to understand the links of these basic spectra to that of the background blackbody spectrum. Both G and Y are generated by applications of the boost generator[5], $\hat{O}_x = -x\partial_x$:

$$G(x) = \hat{O}_x n_{\rm bb}(x), \qquad Y(x) = \hat{\mathcal{D}}_x n_{\rm bb}(x) = \hat{O}_x(\hat{O}_x - 3) n_{\rm bb}(x). \tag{2.10}$$

Making the Ansatz $\Delta n = \Theta\, G(x) + y\, Y(x) + \mu M(x)$ and inserting back into Eq. (2.7) we obtain

$$\Theta' G(x) + y' Y(x) + \mu' M(x) = \Theta_{\rm e}\, Y(y) + \Theta\, \hat{\mathcal{K}}_x\, G(x) + y\, \hat{\mathcal{K}}_x\, Y(x) + \mu\, \hat{\mathcal{K}}_x\, M(x), \tag{2.11}$$

where the prime indicates the derivative with respect to y_z. The stationary solution of the Kompaneets equation, $\hat{\mathcal{K}}_x\, G(x)/x = 0$, defines the null-space. Hence, the number-conserving definition of the μ-distortion, which we are using here, will transform like the temperature shift. In fact both functions

$$K_G \equiv \hat{\mathcal{K}}_x G(x) = -Y(x) \qquad \text{and} \qquad K_M \equiv \hat{\mathcal{K}}_x M(x) = -Y(x)/\beta_M \equiv -\eta_M Y(x) \tag{2.12}$$

nicely map back onto $Y(x)$. However, the function $K_Y(x) = \hat{\mathcal{K}}_x Y(x)$ has contributions that are not spanned by $G(x)$, $Y(x)$ and $M(x)$. We can nevertheless enforce a representation of $K_Y(x)$ in terms of $G(x)$, $Y(x)$ and $M(x)$. Here, we are mostly interested in intensity functions. Knowing that $K_Y(x)$ does not carry photon number,[6] we understand that only $Y(x)$ and $M(x)$ can contribute in our current basis. Since with our finite basis, the representation will not be exact, we can demand that the contribution of M to follow from energy conservation to improve matters, as discussed below.

This operator-function view of the mathematics allows us to make a choice of an inner product on the functions. We can choose to define the matrix elements of the operator $\hat{X}$ between the two function $F(x)$ and $J(x)$ as

$$\langle F|\hat{X}|J\rangle \equiv \langle F|\hat{X}J\rangle = \int x^3 F(x)\, [x^3 \hat{X} J(x)]\, \mathrm{d}x. \tag{2.13}$$

Although by no means rigorous, this inner product-like operation helps us to define the expansion coefficients heuristically. It is equivalent to taking the integrals over the two intensities $x^3 F$ and $x^3 \hat{X} J$, and will be helpful later in ensuring energy conservation of the entire system.

---

[5] This name is related to the fact that Lorentz transformations of the average blackbody spectrum involve this operator.

[6] The integral $\int x^2 K_Y(x)\, \mathrm{d}x = \int x^2 \hat{\mathcal{K}}_x Y(x)\, \mathrm{d}x$ vanishes since the Kompaneets operator conserves photon number.

To decompose $K_Y = \hat{\mathcal{K}}_x Y$ in the most simple approach we remap back to the basis using the Ansatz $|\hat{\mathcal{K}}_x Y\rangle \approx a_0 |Y\rangle + a_1 |M\rangle$ and solve the system

$$\langle Y|\hat{\mathcal{K}}_x|Y\rangle \equiv \langle Y|K_Y\rangle \approx \langle Y|Y\rangle\, a_0 + \langle Y|M\rangle\, a_1 \tag{2.14a}$$

$$\langle M|\hat{\mathcal{K}}_x|Y\rangle \equiv \langle M|K_Y\rangle \approx \langle M|Y\rangle\, a_0 + \langle M|M\rangle\, a_1. \tag{2.14b}$$

This system is equivalent to the matrix equation $\boldsymbol{b} = M_R\, \boldsymbol{a}$, where $b_i = \langle R_i|K_Y\rangle$ for each function of the representation basis, i.e., $R_0 = Y$ and $R_1 = M$ in the considered case. Similarly, we have the basis mixing matrix $M_{R,ij} = \langle R_i|R_j\rangle$ and the corresponding representation coefficients a_0 and a_1. The solution is then $\boldsymbol{a} = {M_R}^{-1}\, \boldsymbol{b}$, such that $K_Y \approx \boldsymbol{R} \cdot \boldsymbol{a}$ with $\boldsymbol{R} = (Y(x), M(x))^T$ and $\boldsymbol{a} = (a_0, a_1)^T$. Carrying out the projection integrals and inverting the system we obtain $K_Y \approx -8.8169\, Y(x) + 40.409\, M(x)$. However, since we used an incomplete basis, this approximation does not satisfy energy conservation. Carrying out the energy integrals, we find $E_{K_Y} = \int x^3 K_Y(x)\, \mathrm{d}x \approx -21.598$ by direct integration of the exact function and $E_{K_Y} \approx -8.8169 \times 4 + 40.409/1.4007 = -6.4178$ from the approximation. Since energy and photon number conservation are the *most fundamental* aspects of the thermalisation problem, this is not a solution we can work with.

To fix the problem, we replace the last equation in the system, Eq. (2.14), with the energy conservation equation. This yields the augmented system

$$\langle Y|K_Y\rangle \approx \langle Y|Y\rangle\, a_0 + \langle Y|M\rangle\, a_1 \tag{2.15a}$$

$$E_{K_Y} = E_Y\, a_0 + E_M\, a_1, \tag{2.15b}$$

which can still be thought of as $\boldsymbol{b} = M_R\, \boldsymbol{a}$, but with modified last rows in $\boldsymbol{b}$ and M_R according to the energy conservation equation. By inverting the new system, this then yields the improved representation $K_Y(x) \approx -3.4593\, Y(x) - 10.871\, M(x)$. Carrying out the energy integrals, we find $E_{K_Y} \approx -3.4593 \times 4 - 10.871/1.4007 = -21.598$, in agreement with the direct integral result.

We have now reformulated the problem once we also determine $\Theta_{\rm e} \approx \Theta_{\rm eq}$. In vector notation, our Ansatz reads $\Delta n = \boldsymbol{B} \cdot \boldsymbol{y}$, where now we include $G(x)$ in the basis, i.e., $\boldsymbol{B} = (G(x), Y(x), M(x))^T$ and $\boldsymbol{y} = (\Theta, y, \mu)^T$. By inserting this Ansatz for Δn into Eq. (2.8) for the Compton equilibrium temperature perturbation, and carrying out the energy exchange integrals one finds

$$\Theta_{\rm eq} \approx \left(\frac{\int x^3 w_y(x)\, \boldsymbol{B}\, \mathrm{d}x}{4E_{n_{\rm bb}}}\right) \cdot \boldsymbol{y} = \eta_G\, \Theta + \eta_Y\, y + \eta_M\, \mu \approx \Theta + 5.3996y + 0.4561\mu. \tag{2.16}$$

Inserting everything back into Eq. (2.11) and collecting terms, with Eq. (2.12) we then obtain

$$\begin{aligned}\Theta' G(x) + y' Y(x) + \mu' M(x) &= \Theta_{\rm e} Y(x) - \Theta\, Y(x) + y\, K_Y - \mu\, \eta_M Y(x) \\ &\approx 1.9403\, y\, Y(x) - 10.871\, y\, M(x).\end{aligned} \tag{2.17}$$

We note that the terms in the Compton equilibrium temperature $\propto \Theta$ and μ cancel identically due to the identities in Eq. (2.12). We furthermore comment that Eq. (2.17) can be also obtained by directly carrying out the projections onto the basis starting from Eq. (2.7). We show this more formally in Appendix B for the extended basis that is discussed in the next section.

Since the system in Eq. (2.17) has to be fulfilled for any x and because the spectral basis is non-degenerate, by comparing coefficients, we obtain the ordinary differential equation (ODE) system

$$\Theta' \approx 0, \qquad y' \approx 1.9403\, y, \qquad \mu' \approx -1.9403\, (\epsilon_Y/\epsilon_M)\, y. \tag{2.18}$$

with $\epsilon_Y/\epsilon_M \approx 5.6026$. While these equations correctly represent the conservation of photon number [only $G(x)$ carries photon number but Θ does not change] and also energy (the sum of the energies in μ and y does not change), they do not yield the correct overall evolution: For the y parameter, the solution is $y(y_z) \simeq y(0)\,\mathrm{e}^{1.9403 y_z}$, while we saw in Sect. (2.2) that it should be more close to $y(y_z) \approx y(0)\,\mathrm{e}^{-4y_z}$, which due to the sign ensures that energy correctly flows from $y \to \mu$. *What has gone wrong?* The approximate representation of $K_Y(x) \approx -3.4593\, Y(x) - 10.871\, M(x)$ is insufficient, as could have been guessed. This can be appreciated in Fig. 2, where we compare the exact solution of $K_Y(x)$ with various approximations. In particular the high-frequency part of $K_Y(x)$ is not well-captured by this simplest approximation, a problem that we fix next.

2.5 Extension of the basis

To make progress, we need to extend the spectral basis, $\boldsymbol{B}$. One of the natural selections is to use the boost operator $\hat{O}_x$ to find the extensions.[7] In principle other choices that could potentially even simplify the calculation either from the theoretical or the computational point of view can be imagined. Our choice is motivated by the fact that $\hat{O}_x$ is one of the fundamental operators generating the Kompaneets operator, $\hat{\mathcal{K}}_x$. In fact, at $x \ll 1$, the Kompaneets operator reduces to $\hat{\mathcal{K}}_x^{\rm low} = 6 + \hat{O}_x(\hat{O}_x - 5)$, which commutes with $\hat{O}_x$ (i.e., $[\hat{\mathcal{K}}_x^{\rm low}, \hat{O}_x] = 0$) indicating a common basis. The boost operator also commutes with the diffusion operator, $[\hat{\mathcal{D}}_x, \hat{O}_x] = 0$, which further supports this choice also in more general cases (see Appendix A). Finally, it appears in log-moment expansions of distortion spectra related to heating processes, which were shown to have useful properties in terms of gauge-choices [32, 34, 64].[8] For $G(x)$, we have $\hat{O}_x G(x) = \hat{O}_x^2 n_{\rm bb}(x) = 3G(x) + Y(x)$, which directly maps back onto the old basis. For the boosts of $M(x)$ and $Y(x)$, new spectral shapes are generated. However, since $K_M = \hat{\mathcal{K}}_x M(x) = -\eta_M Y(x)$ already maps back onto our basis, for now we only need to think about extensions based on the functions

$$Y_k(x) = (1/4)^k \hat{O}_x^k Y(x). \tag{2.19}$$

These functions can be readily computed using `Mathematica` or through the combinatoric sums given in Appendix A, and are illustrated for a few cases in Fig. 3. The Y_k are similar to those functions appearing in asymptotic expansions of the SZ effect [65–68] and can furthermore be found in perturbative expansions of the photon transfer problem [42, 55, 69, 70]. Note that $Y_0(x) \equiv Y(x)$. We also added the factor of $(1/4)^k$ to make each of the Y_k more comparable in amplitude. This choice also ensures $\epsilon_{Y_k} \equiv 4$. These functions all conserve photon number ($\int x^2 Y_k(x)\,\mathrm{d}x = 0$) and hence provide a natural extension of the simple Y and M basis. As we will see in paper II, these also naturally appear once Doppler-driving in the perturbed Universe is included.

2.6 Generalization of the ODE system

In this section, we outline the basic approach for obtaining a generalized ODE system in the extended basis. By deciding about how many Y_k we include in the Ansatz for Δn, we have to determine the representations for each of the[9] $K_{Y_k} = \hat{\mathcal{K}}_x\, Y_k$ within this basis. To simplify the notation, let us again write the extended representation basis as a vector $\boldsymbol{R}(x) = (Y(x), Y_1(x), \dots, Y_N(x), M(x))^T$. We will denote $R_0 = Y_0 \equiv Y$ and $R_{N+1} = M$, with all the other $R_k = Y_k$ in between. We then make the Ansatz $K_{Y_k} \approx \boldsymbol{R} \cdot \boldsymbol{a}_{Y_k}$ with $\boldsymbol{a}_{Y_k}$ denoting the coefficients of each term in the representation basis, $\boldsymbol{a} = (a_0, a_1, \dots, a_N, a_{N+1})^T$. As above, we now have to compute the projection of K_{Y_k} onto each of

---

[7] We stick to the terminology of *boost operator* because it is the generator of infinitesimal Lorentz boosts. In fact, the Kompaneets equation merely follows from Lorentz boosts up to second order in the electrons velocity.

[8] Our considerations hold when working with the background spectrum. The picture will be more complicated in the presence of inhomogeneities (e.g., see paper II).

[9] These arise from applying $\hat{\mathcal{K}}_x$ to the Ansatz for Δn.

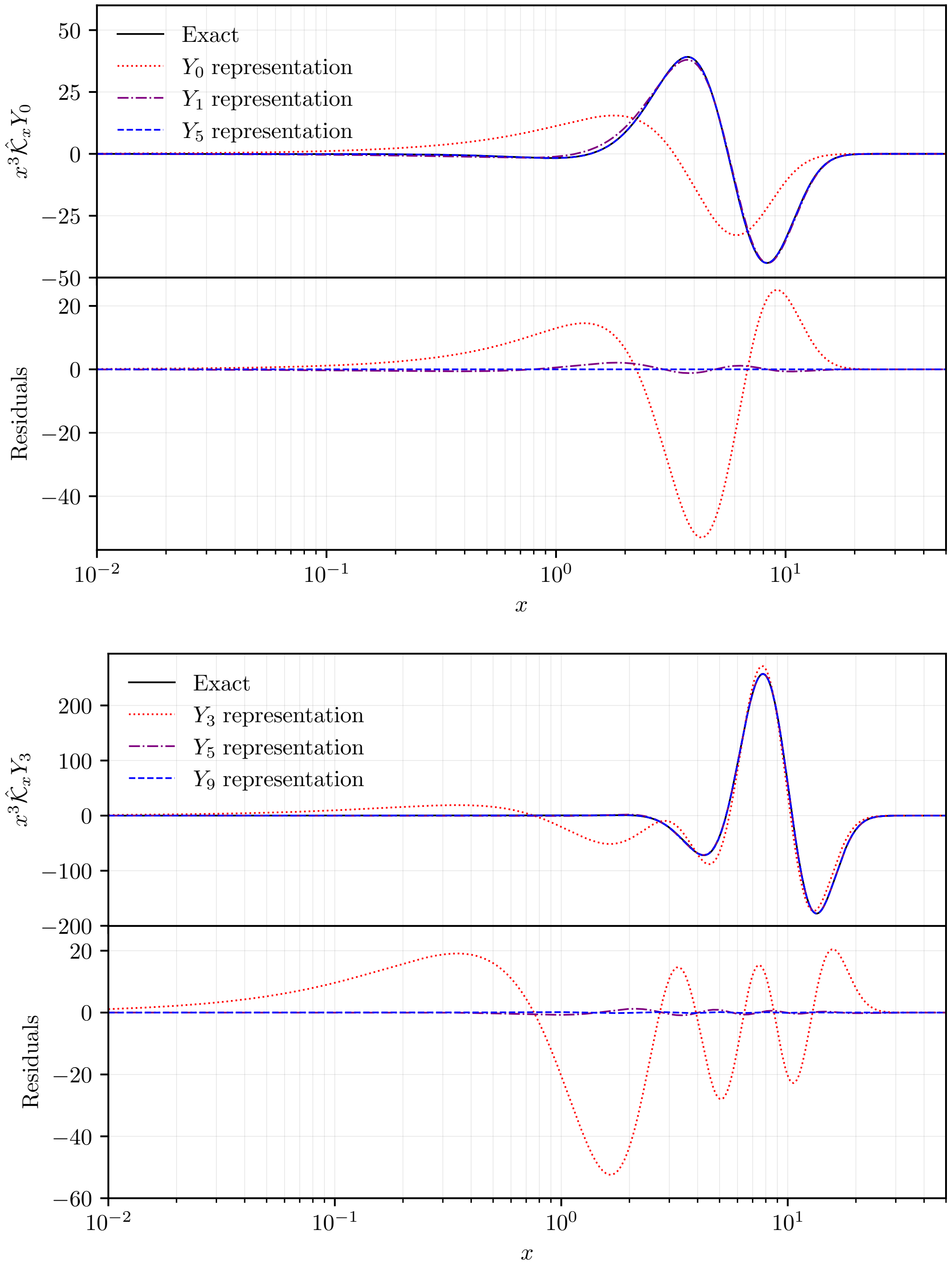


Figure 2: Distortion shapes $x^3 K_Y(x)$ and $x^3 K_{Y_3}(x)$ for various approximations. Here the order refers to the largest term in Y_k that is included [i.e., 0. order only $Y(x)$ and $M(x)$; 5. order includes $M(x)$ and all $Y_k(x)$ up to $Y_5(x)$]. The representations become increasingly accurate the more terms we add to the basis. Typically only poor representation is obtained upon acting on the largest function in the basis – a problem which is mitigated by the fact that less energy occupy these higher modes in numerical solutions.

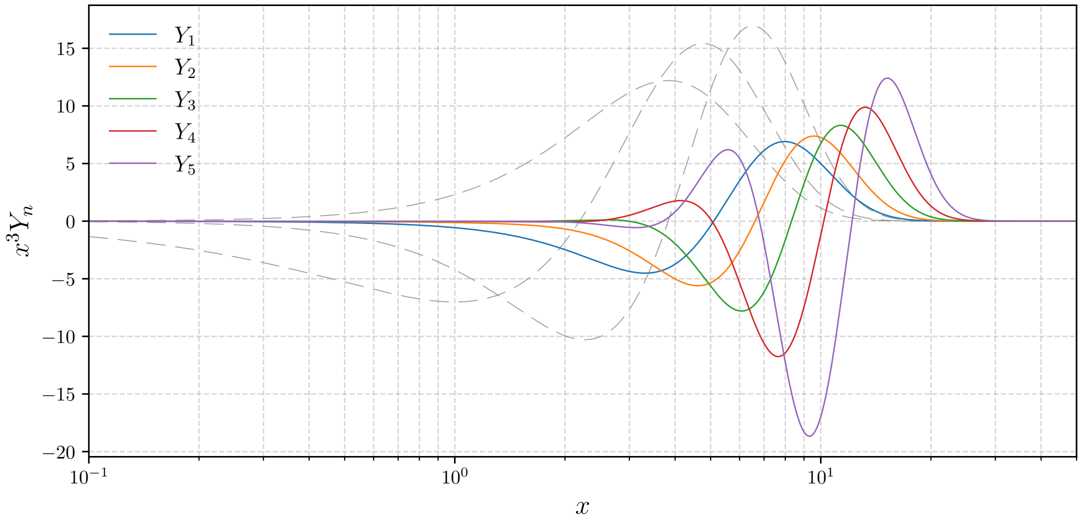


Figure 3: First few basis function Y_k in comparison with the standard distortion shapes (dashed lines, with the positive peak from left to the right relating to G, M, Y, respectively). For increasing k the functions Y_k occupy more and more of the high frequency part of the spectrum.

the R_i. To ensure energy conservation, we will again determine μ using the energy integrals.[10] This then yields the following system of equations that determines the representation vector $\boldsymbol{a}_{Y_k}$:

$$\begin{aligned}
\langle Y|K_{Y_k}\rangle &\approx \langle Y|Y\rangle\, a_{Y_k,0} + \langle Y|Y_1\rangle\, a_{Y_k,1} + \ldots + \langle Y|Y_{N-1}\rangle\, a_{Y_k,N-1} + \langle Y|Y_N\rangle\, a_{Y_k,N} + \langle Y|M\rangle\, a_{Y_k,N+1}\\
\langle Y_1|K_{Y_k}\rangle &\approx \langle Y_1|Y\rangle\, a_{Y_k,0} + \langle Y_1|Y_1\rangle\, a_{Y_k,1} + \ldots + \langle Y_1|Y_{N-1}\rangle\, a_{Y_k,N-1} + \langle Y_1|Y_N\rangle\, a_{Y_k,N} + \langle Y_1|M\rangle\, a_{Y_k,N+1}\\
\vdots\quad &\approx \quad\vdots\\
\langle Y_N|K_{Y_k}\rangle &\approx \langle Y_N|Y\rangle\, a_{Y_k,0} + \langle Y_N|Y_1\rangle\, a_{Y_k,1} + \ldots + \langle Y_N|Y_{N-1}\rangle\, a_{Y_k,N-1} + \langle Y_N|Y_N\rangle\, a_{Y_k,N} + \langle Y_N|M\rangle\, a_{Y_k,N+1}\\
E_{Y_k} &\approx E_Y\, a_{Y_k,0} + E_{Y_1}\, a_{Y_k,1} + \ldots + E_{Y_{N-1}}\, a_{Y_k,N-1} + E_{Y_N}\, a_{Y_k,N} + E_M\, a_{Y_k,N+1},
\end{aligned} \tag{2.20}$$

The last equation is the energy conservation equation to determine the coefficient of $M(x)$. We thus have a matrix equation of the form $\boldsymbol{b}_{K_{Y_k}} = M_R\, \boldsymbol{a}_{K_{Y_k}}$, which we can solve for $\boldsymbol{a}_{K_{Y_k}}$ given a finite representation basis. We again highlight the fact that the system was obtained using the energy conservation equation. The matrix M_R is therefore again nearly equivalent to the full basis mixing matrix $M_{R,ij} = \langle R_i|R_j\rangle$. However, the last equation is replaced by the energy conservation equation, even if not explicitly distinguished in the notation.

As an example, if we choose $\boldsymbol{R}(x) = (Y(x), Y_1(x), M(x))^T$, we only have to determine the representations for $\hat{\mathcal{K}}_x Y$ and $\hat{\mathcal{K}}_x Y_1$. Solving the corresponding systems of equations then yields

$$K_Y(x) \approx 2.4717\, Y(x) - 8.4907\, Y_1(x) + 3.4698\, M(x) \equiv \boldsymbol{R}\cdot\boldsymbol{a}_Y \tag{2.21a}$$

$$K_{Y_1}(x) \approx 28.134\, Y(x) - 26.125\, Y_1(x) - 55.089\, M(x) \equiv \boldsymbol{R}\cdot\boldsymbol{a}_{Y_1}. \tag{2.21b}$$

Looking at Fig. 2, we see that now the match with the exact result for $K_Y(x)$ is already very good. We note that adding more terms to the basis evidently changes *all* the coefficients of the solution, and also improves the result for the correspondence (see Fig. 2).

---

[10] We could really replace any one equation using energy conservation.

Inserting Eq. (2.21) back into Eq. (2.7) and using $\eta_{Y_1} \approx 7.8246$ in Eq. (2.8), after collecting coefficients we then find

$$\begin{aligned}\Theta' G(x) + y' Y(x) + y_1' Y_1(x) + \mu' M(x) &= \Theta_{\rm e}\, Y(x) - \Theta\, Y(x) + y\, K_Y(x) + y_1\, K_{Y_1}(x) - \mu\, \eta_M Y(x) \\ &\approx (7.8714y + 35.958y_1)\, Y(x) - (8.4907y + 26.125y_1)\, Y_1(x) + (3.4698y - 55.089y_1)\, M(x).\end{aligned} \tag{2.22}$$

In Appendix B we give an alternative derivation that avoids the intermediate step of first representing the K_{Y_k} in terms of the basis. However, mathematically this is equivalent. By comparing the coefficients, one can again obtain a system for the evolution of Θ, y, y_1 and μ. The solution of this system now has the correct main properties. It conserves photon number and energy and leads to a solution[11] $y(y_z) \simeq y(0)\, {\rm e}^{-3.8\, y_z}$. Indeed this is very close to the correct Green's function solution that neglects any residual distortion contributions. However, the precision can be improved by further extending the spectral basis (see Fig. 2).

Below we will give the solutions for systems that include up to $Y_{15}(x)$ in the basis. This already provides a very accurate approximation for the exact themalisation Green's function. The related system can be readily generated using `Mathematica` following the procedure above. Schematically, we can then express the effect of the Kompaneets operator on the distortion in the form

$$\Delta n' = \Theta_{\rm e}\, Y + \hat{\mathcal{K}}_x\, \Delta n \qquad \longleftrightarrow \qquad \boldsymbol{y}' \approx M_{\rm K}\, \boldsymbol{y} \tag{2.23}$$

with $\boldsymbol{y} = (\Theta, y, y_1, \dots, y_N, \mu)^T$ and where $M_{\rm K}$ is the Kompaneets mixing matrix that directly depends on the chosen spectral basis.[12] By construction, this matrix merely rotates $y \to y_k \to \mu$ under energy conservation and thus has a zero first row and column. Even order systems are omitted, as they are found to be numerically unstable.[13] We suspect this is due to the second order nature of the Kompaneets operator, but have no additional prove for this. The solution at any moment is then $\Delta n(x, y_z) \approx \boldsymbol{B}(x) \cdot \boldsymbol{y}(y_z)$ with the full spectral basis $\boldsymbol{B}(x) = (G(x), Y(x), Y_1(x), \dots, Y_N(x), M(x))^T$.

2.7 Adding the effect of photon production and heating

To add the effect of photon production by double Compton (DC) and Bremsstrahlung (BR), we make use of the fact that once these become important, the y_k will be extremely short-lived (i.e., decay quickly, $y_k \to 0$). In this case, we can neglect the role of the Y_k's for photon production and the analytic results for the μ-distortion evolution can be used [1, 61]. The net photon emission and absorption term has the explicit form [7, 58, 61]

$$\frac{1}{\dot\tau}\,\frac{\partial n_0}{\partial t}\bigg|_{\rm em/abs} = \frac{\Lambda(x, \theta_{\rm e}, \theta_z)\, {\rm e}^{-x\,\theta_z/\theta_{\rm e}}}{x^3}\left[1 - n_0\left({\rm e}^{x\,\theta_z/\theta_{\rm e}} - 1\right)\right] \approx -\frac{\Lambda(x, \theta_z)\,(1 - {\rm e}^{-x})}{x^3}\Delta n_0 + \frac{\Lambda(x, \theta_z)}{x^2}\, n_{\rm bb}\, \Theta_{\rm e}.$$

In the last step, we again linearised the problem with respect to the distortion [and $\Theta_{\rm e} \simeq \mathcal{O}(\Delta n)$]. The DC and BR emissivities can be computed accurately using `DCpack` [71] and `BRpack` [72].

As already explained in Sect. 2.2, we can think of the effect that photon emission and absorption has on the distortion as a *redistribution* between μ and Θ. Overall this means

$$\frac{\partial n_0}{\partial y_z}\bigg|_{\rm em/abs} \longleftrightarrow \gamma_T\, x_{\rm c}\, \mu\, G(x) - \gamma_N\, x_{\rm c}\, \mu\, M(x), \tag{2.24}$$

---

[11] This can be seen when assuming that the coefficient of Y_1 evolves under quasi-stationary conditions. This implies the condition $8.4907y + 26.125y_1 \approx 0$ resulting in $y_1^{\rm qs} \approx -0.32550\, y$, which yields the desired result similar to Eq. (2.4).

[12] We will provide the system for up to y_{15} under `www.Chluba.de/CosmoTherm`.

[13] One can change the weight function in the definition of the scalar-product, Eq. (2.13), to remedy this issue, but we did not explore this option any further.

as in Eq. (2.5). This greatly simplifies the thermalisation problem, essentially converting the collision term into a *source-sink* term with built-in energy conservation.

To also add the effect of external heating, we assume that the distortions are generated through a y-distortion source, $y' = (1/4)\,Q'$, where $Q' = \mathrm{d}Q/\,\mathrm{d}y_z = (\dot{\tau}\theta_z)^{-1}\,\mathrm{d}Q/\,\mathrm{d}t$ in this context. For energy release scenarios, this will be a very good approximation in the pre-recombination era, since heat that is transferred to the baryons quickly reaches the photons through Compton scattering [e.g., see 7, 59]. The factor of $\alpha_Y = 1/4$ converts the change of the relative energy density into the y-parameter. Together we then have

$$\begin{aligned} \Delta n' = \Theta_\mathrm{e}\, Y + \hat{\mathcal{K}}_x\, \Delta n + \Delta n'|_{\mathrm{em/abs}} + \Delta n'|_\mathrm{h} \quad &\longleftrightarrow \quad \boldsymbol{y}' \approx M_\mathrm{K}\, \boldsymbol{y} + \boldsymbol{D} + \frac{\boldsymbol{Q}'}{4}, \\ \boldsymbol{D} = (\gamma_T x_\mathrm{c}\, \mu, 0, 0, \dots, 0, -\gamma_N x_\mathrm{c}\, \mu)^T\,, \quad &\boldsymbol{Q}' = (0, Q', 0, \dots, 0, 0)^T\,. \end{aligned} \tag{2.25}$$

This equation now allows us to account for the effects of external heating and emission/absorption with the source vectors, $\boldsymbol{Q}'$ and $\boldsymbol{D}$, respectively. Refinements to the treatment of photon emission and absorption that include the effects of $Y_k(x)$ as well as other corrections to the leading order terms can in principle be added following the method of [61]; however, for now we stop with this simple description, emphasizing again that most of the thermalisation of distortions occurs deep into the μ-era, when these effects are expected to be small.

It is worth noting that by formulating the heating as a vector in the extended spectral basis this assumption can be generalised and other spectral shapes can in principle be sourced at late times [see 73, for photon injection distortion evolution]. This will not be explored in this paper – we adopt the standard thermalisation picture and study excitations of the y-distortion amplitude through energy injection as our benchmark in the following sections.

2.8 Solutions for the Green's function after single injection

The thermalisation Green's function has been successfully used to represent the spectral distortion shapes from continuous heating [50, 53, 74]. With the above description we can reproduce the Green's function to high precision, as we show now. For this, we model the scenario of single injection in Eq. (2.25) and introduce a narrow Gaussian heating rate at $z_\mathrm{injection}$ (or alternatively set an effective initial condition for y at that redshift). Allowing this to evolve under successive scatterings we study the state of the system $\boldsymbol{y}(z_\mathrm{f})$ at the final redshift z_f, and extract the corresponding spectral shape.

The result of this calculation for various basis sizes is shown in Fig. 4. Also shown is a comparison to the exact result of `CosmoTherm`, which performs an analogous calculation by directly binning the frequency space. The latter approach can be thought of as applying a "top hat" basis in x to the same formalism discussed in Sect. 2, and thus is more precise at the cost of tracking thousands of equations simultaneously. Despite the relative simplicity of the treatment derived here it is possible to capture the transition from temperature shift to y-distortion through the intermediate μ and residual eras accurately. The residual era in particular is captured by the expanded basis Y_k, with $N_\mathrm{max} = 9$ already yielding very accurate results.

It is noteworthy that with the inclusion of the new spectral shapes Y_k the definitions of y and μ have a degree of degeneracy. This is most notable in the recession of the μ-era with increasing N_max and the existence of $\mathcal{J}_y > 1$ in the residual era (energy conservation is ensured by cancellation with the negative $\mathcal{J}_{y_k}$). This apparent *arbitrary* labelling of energy with different coefficients is not problematic, since the real physical observable that must converge is the spectrum, which indeed remains stable as seen in the right panels of Fig. 4. This physical observable will itself be projected onto some beneficial spectral shapes, as discussed in Sect. 3, which depend on the characteristics of observing instrument [10, 50] or other theoretical choices.

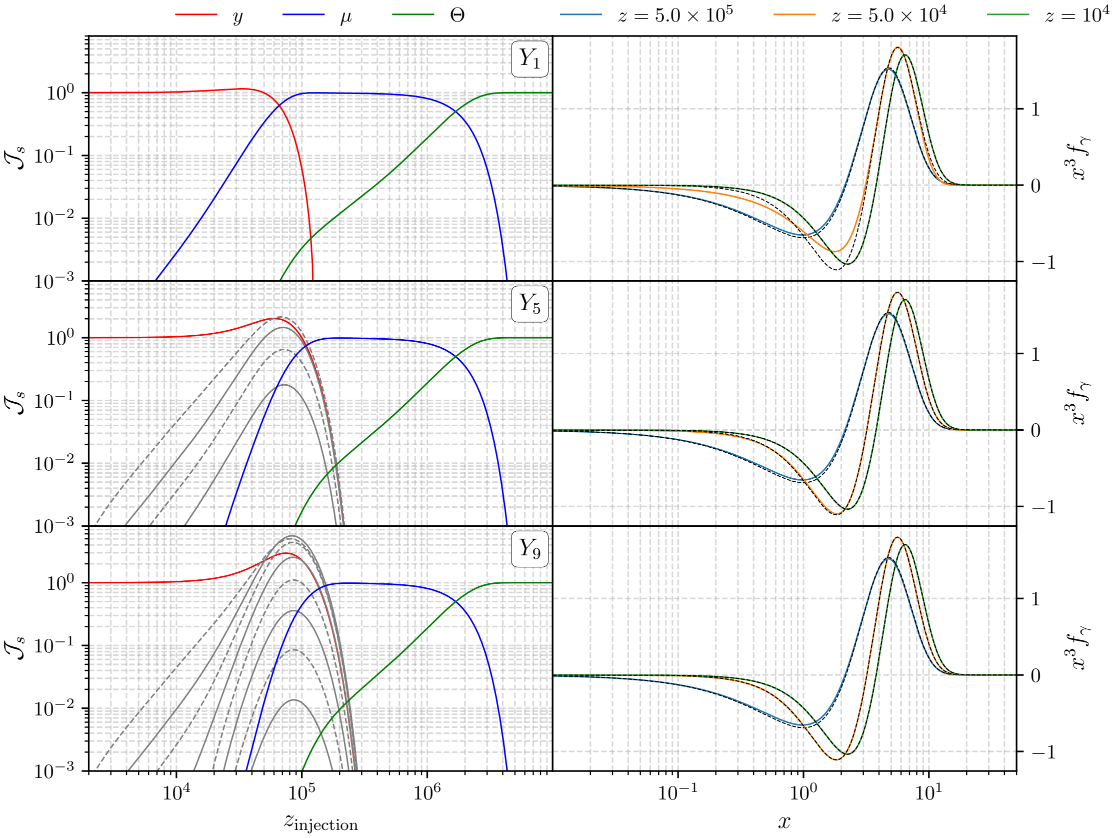


Figure 4: A figure showing the iterative improvements of augmenting the Y_k basis. The rows show the branching ratios across redshifts (left) and final spectrum at various energy injection redshifts (right) for $N_{\rm max} = 1, 5, 9$. Gray lines in the branching ratio plots correspond to the $Y_{k>0}$ coefficients, and dotted lines are negative values. Dotted black lines in the spectrum plots show the full results performed with `CosmoTherm`.

While the spectra in Fig. 4 show each snapshot being captured accurately, we note that the precise timings of the transition from one phase to the next appear slightly delayed relative to the full `CosmoTherm` calculation.[14] In Fig. 5 we show two time slices in the transition phases $T \rightarrow \mu$ and $\mu \rightarrow y$, again with their respective `CosmoTherm` comparison and an optimised least squares fit to the full solution using the approximate basis of spectral functions. At $z = 5 \times 10^5$ (left panel) we see the approximate solution having part of a temperature shift while the full solution is almost a pure μ distortion. Seeing that the optimised fit reproduces the `CosmoTherm` solution well, we conclude that the approximate treatment slightly overestimates the thermalisation timescale. As explained in [61], several additional aspects that are not captured by the simple treatment here do matter at the level of a few percent. By more carefully treating the DC and BR thermalisation rate, which will lead to a refined scaling of $x_{\rm c}$ with time, one can probably improve the treatment; however, for our purpose the current approximation shall suffice, and refinements are left to future work.

In the right panel of Fig. 5, we also see a snapshot at $z = 10^5$. It is apparent that the approximate basis approaches the full solution, but does not capture it fully. For comparison, an optimised fit (which allows one to smooth over any time-dependent mismatch) is shown but also fails to exactly reproduce the curve in this case. We can therefore conclude the basis only has enough freedom to

---

[14] Videos illustrating the solution will become available at `www.Chluba.de/CosmoTherm`.

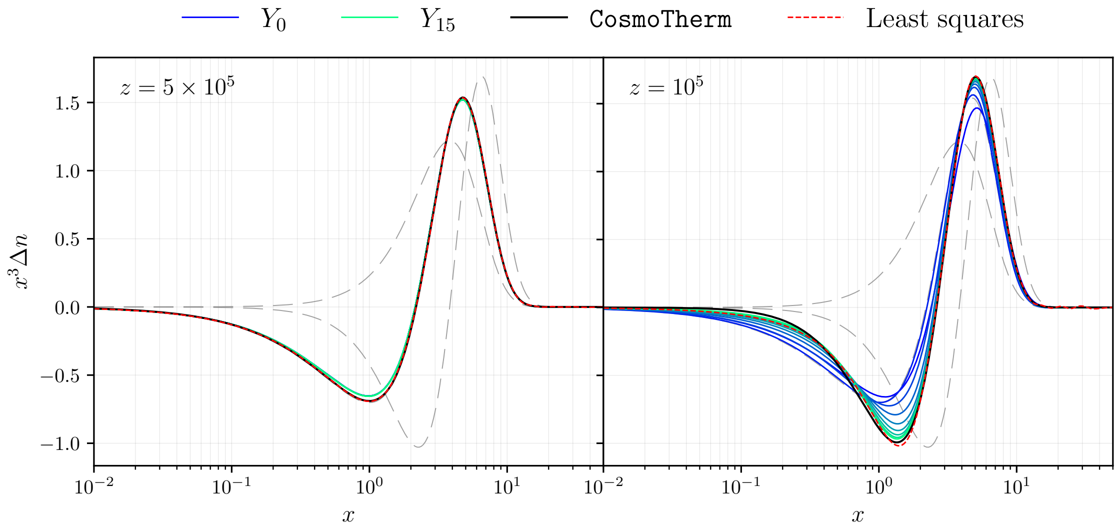


Figure 5: A figure showing two characteristic time slices where the full solution is poorly approximated. At $z = 5 \times 10^5$ (left) the approximate basis still has an excessive contribution from G. At $z = 10^5$ (right) the different order of bases approach the residual distortion shape, but perform especially badly at low-frequencies. Also shown is a least-squares fit (red dotted line) using the Y_{11} basis.

capture some — but not all — of the nuances of the residual era. Departures from the solution are visible at low and intermediate frequencies (i.e., $x \lesssim 2$) owing to the nature of the chosen basis (see Fig. 3) and our focus on energy conservation, which is driven by the high-frequency tail. Additional work on the optimal basis will likely remedy these limitations; however, we highlight that our treatment already greatly improves the modeling of the residual era, which is barely captured using a simple $y + \mu$ approximation. Hence, we again shall be content with the performance of the ODE treatment and focus on applications to anisotropic distortions as the main next step (paper II).

3 Defining spectral distortion observables

In Fig. 4 we saw that the amplitude of the y distortion changes depending on the other amplitudes within the expanded basis while leaving the actual photon spectrum unchanged. In this section, we will formalize and further discuss this phenomenon in the context of changing basis. Heuristically, we can see the space of valid spectra as an abstract vector space, and as such choose a basis for this space. Provided the spectrum is a continuous function we expect a formal basis to be infinite, but computationally a finite basis can suffice, if chosen well. The bottom line statement we emphasise and highlight here then is that the underlying physics will be (and must be) independent of the choice of basis, where the physics here is captured only by the full photon spectrum and not any individual branching ratio or transfer function.

This gives the freedom to choose a basis which suits a given purpose most appropriately. In this light we will introduce two new bases, which are useful for packaging and exporting the results of the Y_k basis or *computation basis*. At a given spectral sensitivity, only a finite number of spectral parameters will be directly measurable, and it is moot to attempt determining the amplitudes (or power and cross-power spectra) for all the spectral parameters inherent to the computation basis. The other two bases introduced here are guided by the principle to compress the information in the spectrum and prepare for easily extracting and interpreting the physics in observations.

Although from our discussion it is clear that a better computation basis which captures all the spectral complexity at low frequencies may exist, we are now interested in finding alternative representations for the space spanned by our Y_k basis. As explained in [50], for a given experimental setting (e.g., frequency coverage and channel sensitivities) one can ask which spectral shapes are best constrained aside from the standard distortions. These spectral shapes can be determined using a principal component analysis. Mathematically, this can be thought of as an expansion of the spectrum into μ, y and Θ plus some additional spectral parameters, r_i, to describe the residual distortion shapes.[15] The residual distortion shapes are the principal spectral components spanning the residual distortion space and can be ranked by their observability, defining the *observation basis*. Denoting the residual distortion eigenspectra as $\boldsymbol{S}^{(k)}$, we find

$$\Delta I_i = \int B_i(\nu)\,\Delta I_\nu\,\mathrm{d}\nu = \Theta\,\Delta I_i^G + y\,\Delta I_i^Y + \mu\,\Delta I_i^M + \Delta R_i, \qquad \text{and} \qquad \Delta R_i = \sum_{k=1} r_k\,S_i^{(k)} \tag{3.1}$$

where $\Delta I_\nu = 2h\nu^3/c^2\Delta n_\nu$ is the intensity corresponding to Δn_ν, which is integrated over the bandpass, $B_i(\nu)$. In our computation we shall use a simple top-hat bandpass centered around frequency ν_i with a width $\Delta\nu_i$. Similarly, ΔI_i^G, ΔI_i^Y and ΔI_i^μ are the band-averaged versions of the corresponding G, Y and M intensities. The (band-averaged) residual distortion, ΔR_i, space is orthogonal to M, Y and G, for the selected instrumental configuration. Since the binned spectral shapes can all be thought of as simple vectors, we can directly obtain the μ, y, Θ and r_k values for any distortion signal as

$$\begin{pmatrix} \Theta_\mathrm{o} \\ y_\mathrm{o} \\ \mu_\mathrm{o} \end{pmatrix} = \begin{pmatrix} \Delta\boldsymbol{I}^G\cdot\Delta\boldsymbol{I}^G & \Delta\boldsymbol{I}^G\cdot\Delta\boldsymbol{I}^Y & \Delta\boldsymbol{I}^G\cdot\Delta\boldsymbol{I}^M \\ \Delta\boldsymbol{I}^Y\cdot\Delta\boldsymbol{I}^G & \Delta\boldsymbol{I}^Y\cdot\Delta\boldsymbol{I}^Y & \Delta\boldsymbol{I}^Y\cdot\Delta\boldsymbol{I}^M \\ \Delta\boldsymbol{I}^M\cdot\Delta\boldsymbol{I}^G & \Delta\boldsymbol{I}^M\cdot\Delta\boldsymbol{I}^Y & \Delta\boldsymbol{I}^M\cdot\Delta\boldsymbol{I}^M \end{pmatrix}^{-1} \begin{pmatrix} \Delta\boldsymbol{I}^G\cdot\Delta\boldsymbol{I} \\ \Delta\boldsymbol{I}^Y\cdot\Delta\boldsymbol{I} \\ \Delta\boldsymbol{I}^M\cdot\Delta\boldsymbol{I} \end{pmatrix} \qquad \text{and} \qquad r_k = \frac{\boldsymbol{S}^{(k)}\cdot\Delta\boldsymbol{I}}{\boldsymbol{S}^{(k)}\cdot\boldsymbol{S}^{(k)}}. \tag{3.2}$$

This assumes that the covariance of the spectral bands is diagonal, but extensions can be readily given. The residual distortion parameters, by construction, will *only* receive contributions from the y_i of our computation basis, while Θ_o, μ_o and y_o will be a superposition of the Θ, y and μ values in the previous basis with extra contributions from the y_i. The relevant *rotation* of the basis can be precomputed (see Sect. 3.1). Given the observation basis $\boldsymbol{S}^{(k)}$ we can therefore usually compress the information into fewer observational parameters, as we show below.

In Fig. 6, we show the first few $\boldsymbol{S}^{(k)}$ used in our computations below. The basis was created assuming constant channel sensitivity and channel widths $\Delta\nu = 1\,\mathrm{GHz}$ in the range $\nu_\mathrm{min} = 30\,\mathrm{GHz}$ to $\nu_\mathrm{max} = 1000\,\mathrm{GHz}$, mainly for illustration. We normalized all of these to carry $\Delta\rho_\gamma/\rho_\gamma = 4$ of energy. This choice makes them comparable in amplitude to the standard distortion shapes and the level of the corresponding residual distortion parameter gives away its relative importance. Creating the optimal distortion eigenmodes for more realistic experimental configurations is straightforward following the procedure outlined in [10, 50]. We can see that the distortion eigenmodes exhibit an increasing number of nodes, reminiscent of other orthogonal functions sets. In applications, this will typically lead to the corresponding residual distortion parameter, r_i, decreasing in amplitude.

In Fig. 7 we illustrate how this mapping to the observation basis modifies the appearance of the branching ratios (left panels) and photon spectra (right panels). We immediately see that the y distortion does not take on a relative energy contribution > 1 as it did in the computation basis (Fig. 4). Further to this point, the total amplitude of y and μ are more stable for increasing $y_{n>0}$, revealing

---

[15] In [50] the residual distortion amplitudes are referred to as μ_i, but we shall use a new nomenclature henceforth.

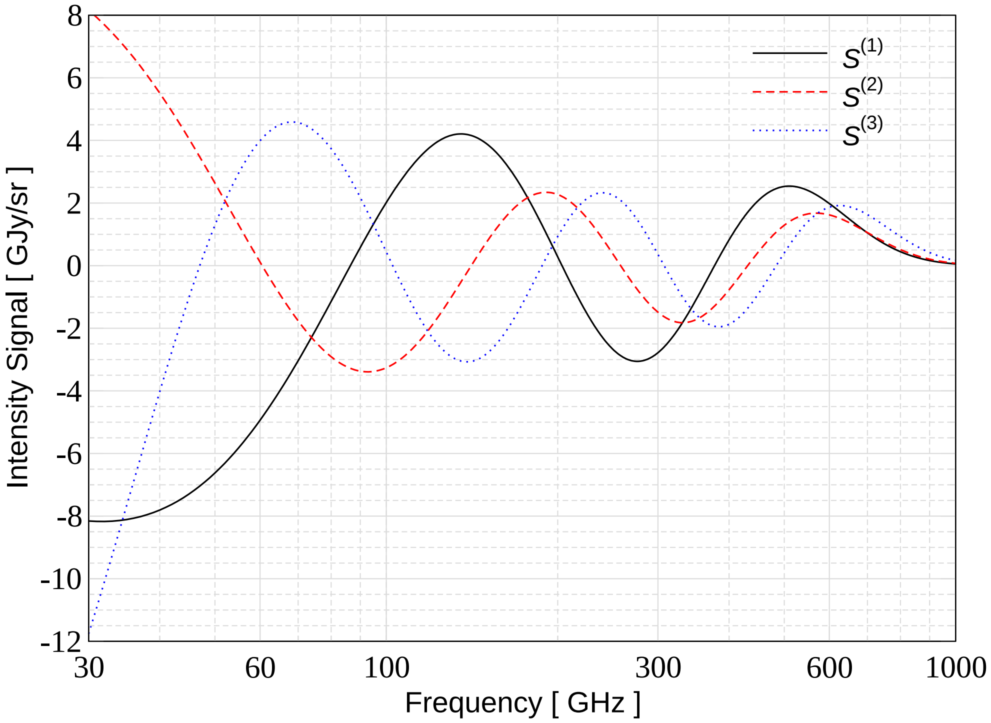


Figure 6: First three residual distortion eigenmodes obtained for $\Delta\nu = 1\,\text{GHz}$ in the range $\nu_{\rm min} = 30\,\text{GHz}$ to $\nu_{\rm max} = 1000\,\text{GHz}$. These signals are orthogonal to the standard G, Y and M spectra and also among each other. They have all been normalized to carry an energy of $\Delta\rho_\gamma/\rho_\gamma = 4$.

that the modelling of spectral evolution is improving with basis size, but without the usual incurred *coefficient ambiguity* as a trade-off. The spectra cover a smaller frequency range, as discussed above, but otherwise show no significant departure from the result of the computational basis. Recall the statement that the bottom line physical result – the spectrum – is independent of the chosen basis.

3.1 Efficient change of the basis

To accelerate the calculation we can precompute all 'rotations' from one basis to the other given the distortion vectors (which depend on the experimental setting). Algorithmically, we have to bin all the involved spectra from the various bases and then compute the relevant mixing matrices and subsequently invert the problem. This then defines the mixing matrix L, which maps $\boldsymbol{y} = (\Theta, y, y_1, ..., y_N, \mu)$ to $\boldsymbol{o} = (\Theta_{\rm o}, y_{\rm o}, r_1, ..., r_M, \mu_{\rm o})$ as $\boldsymbol{o} = L\,\boldsymbol{y}$. The dimension of the two spaces need not be the same, with the observation basis having a lower dimension given that the observability of various independent signal modes is usually reduced.

For our analysis, we pre-compute L for $N = 15$ and $M = 6$, but usually will only need r_1, r_2 and r_3 to obtain a highly accurate representation of the full Y_k basis result. Even for computations of power spectra, this significantly reduces the dimensionality of the problem (see paper III). As shown in Fig. 8, the residual distortion representation performs as well as the computation basis but with a lot fewer components (see discussion next Section).

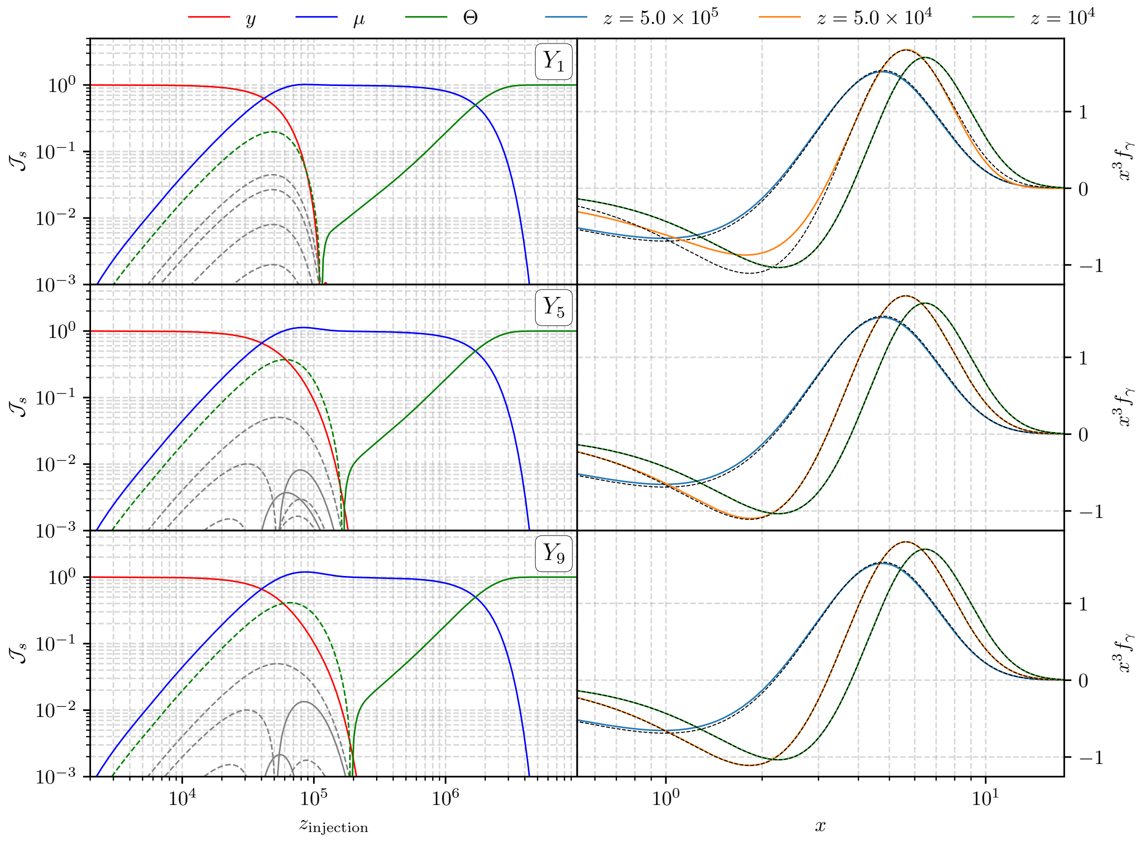


Figure 7: As for Fig. 4, but now with results cast into the "observation" basis. Notice how now the y and μ amplitudes are stable with increasing basis size. The spectra cover a smaller frequency range, as dictated by realistic observational scenarios, however they do not otherwise change compared to the computational basis. The residual era shows an effective negative temperature shift to achieve the correct spectral shape.

3.2 Performance and convergence

We are now in the position to compare the performance of the *observation basis* in representing the distortion solutions obtained by using the computational Y_k basis. Two aspects are immediately worth noting: since the observation basis has a limited frequency coverage, it will not provide a description of the distortion solution *outside* this domain. This is analogous to having limited sky coverage, although there the properties of the spherical harmonic basis allows for some level of statistical deconvolution in CMB analyses [75]. For the distortion spectra, this inversion problem will not be possible unless as many distortion parameters as basis parameters are observed accurately.

Second, the number of independently observable modes will depend on the frequency domain and frequency resolution as well as the sensitivity of the experiment. For example, it has been demonstrated that distinguishing μ-type distortion spectra benefits from having frequency channels below $\simeq 30\,$GHz [44, 76, 77]. However, a more comprehensive exploration of these dependencies aspects is beyond the scope of this paper, and for our illustrations we will stick to the modes shown in Fig. 6.

To illustrate the performance of the observation basis, we consider the distortion caused by a single energy injection at $z_{\rm h} = 5 \times 10^4$ with $\Delta\rho_\gamma/\rho_\gamma = 10^{-5}$. In this regime, the residual distortion contributions are expected to be largest and hence the departures from the standard μ and y description are maximized. Looking at Fig. 8, we can immediately see that only the first few residual distortion

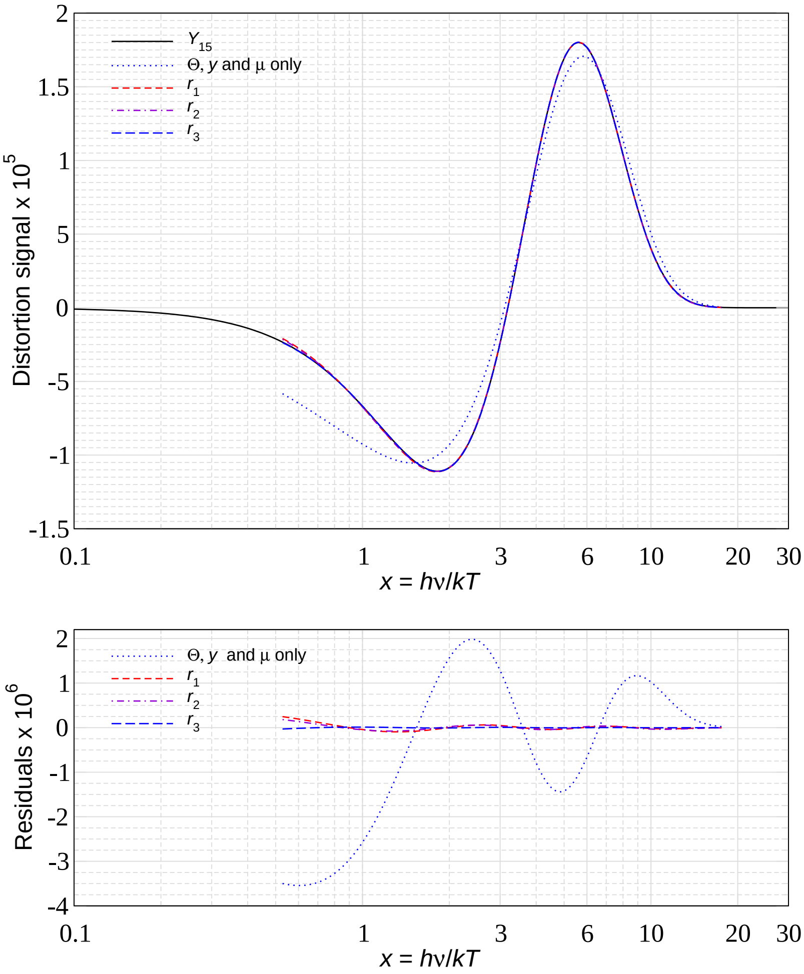


Figure 8: Distortion (i.e., $x^3\Delta n_x$) after a single injection at $z_\mathrm{h} = 5 \times 10^4$ with $\Delta\rho_\gamma/\rho_\gamma = 10^{-5}$ for various representations of the signal. The labels gives the maximal spectral component in the respective basis aside from the standard Θ, y and μ-description. The 'exact' result was obtained with the computation basis up to Y_{15}. The simple Θ, y and μ-descriptions fails at the level of several tens of percent in particular at low frequencies. On the other hand, the distortion is extremely well represented once r_2 or r_3 are included. This is a compression of the information by a factor of more than $\simeq 5$.

spectra are needed to accurately represent the distortion shape at the level of a few percent of the dominant signal. This is a significant compression of the required information for the signal processing. However, it also implies that from the precise distortion shape not as much information can be directly extracted unless a very large distortion signal is present or extremely high sensitivity is achieved [50].

3.3 Caveats of the observation basis and alternative description

We would like to highlight a few important aspects of the observation basis. While by construction, Y, M and the Y_k spectra do not carry photon number, the same is not true for the residual distortion spectra. This implies that in the new representation, not only the y and μ parameters change but also the temperature parameter is affected. Concretely, we have $\Theta \approx 2.6 \times 10^{-12}$, $y \approx 6.1 \times 10^{-6}$ and $\mu \approx 3.0 \times 10^{-8}$ for the example shown in Fig. 8 in the Y_{15}-representation. When projecting onto the

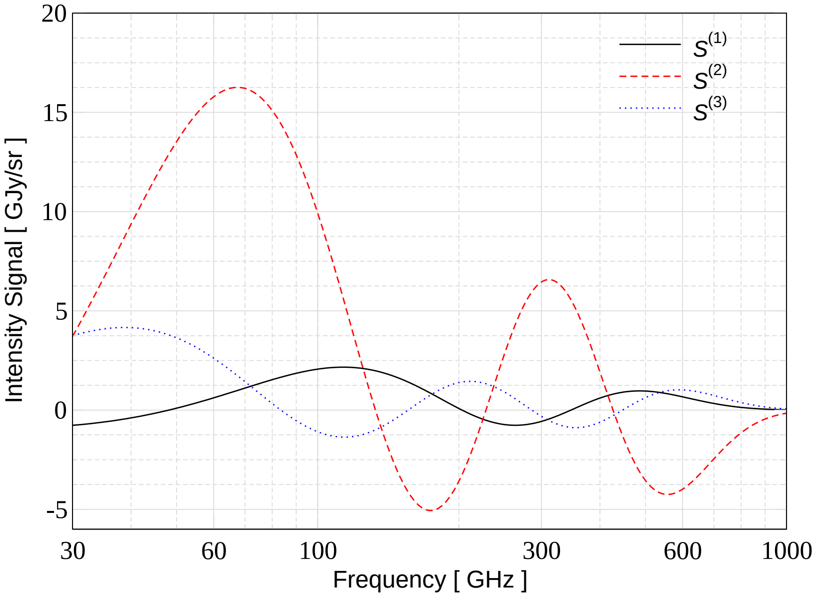


Figure 9: First three residual distortion eigenmodes obtained for $\Delta\nu = 1\,$GHz in the range $\nu_{\rm min} = 30\,$GHz to $\nu_{\rm max} = 1000\,$GHz and with an explicit photon number constraint to deproject G. These signals are orthogonal to the standard Y and M spectra and also among each other, but no longer leave the amplitudes of G unaltered. They have all been normalized to carry an energy of $\Delta\rho_\gamma/\rho_\gamma = 4$.

observation basis we find $\Theta_{\rm o} \approx -9.2\times10^{-7}$, $y_{\rm o} \approx 1.3\times10^{-6}$ and $\mu_{\rm o} \approx 1.3\times10^{-5}$. We injected energy at a redshift where DC and BR are already very inefficient, such that Θ based on scattering physics alone should be negligibly small. After the change of basis, in particular $\mu_{\rm o}$ picks up a noticeable contribution and $\Theta_{\rm o}$ even drops below zero. This effect is known and originates from the fact that the residual distortion construction is based on intensity projections [see Fig. 2 of 50]. The chosen procedure is most close to what would be obtained using standard component separation methods in future spectrometer analyses [e.g., 43, 44, 77]. Although the total energetics of the problem and also the spectrum remain unchanged by the change of the representation, this behavior seems ambiguous.

An alternative observational procedure, without this apparent ambiguity, could be to fix the temperature parameter Θ based on the number density of the photon field. In this case, one could fully orthogonalize G to the distortion space and construct a pure residual *scattering* distortion representation that is unaffected by the aforementioned effects. In Fig. 9, we show the result for the basis vectors in this alternative construction procedure. While generally very similar to the previous set of distortion modes (see Fig. 6), the alternative modes show a slightly differing pattern and overall trend. These modes can only be use in cases where the temperature contribution can be independently separated, as the modes no longer are orthogonal to $G(x)$.

Fig. 10 shows how this basis again stabilises the y and μ amplitudes across basis size while reproducing the same spectrum as the other bases. The difference however is that now the first residual mode constitutes a more dominant fraction of the energy, and the temperature shift never takes on its effective negative value. This is closer to the full scattering physics, since now number is conserved, but the description is not akin to a realistic observation of the sky. In paper III we will use these two bases wherever they are most illustrative, but always being careful and explicit.

In Fig. 11, we demonstrate that these alternative modes also represent the distortion shape very well, with the difference being that the contribution from G was fixed independently using a number

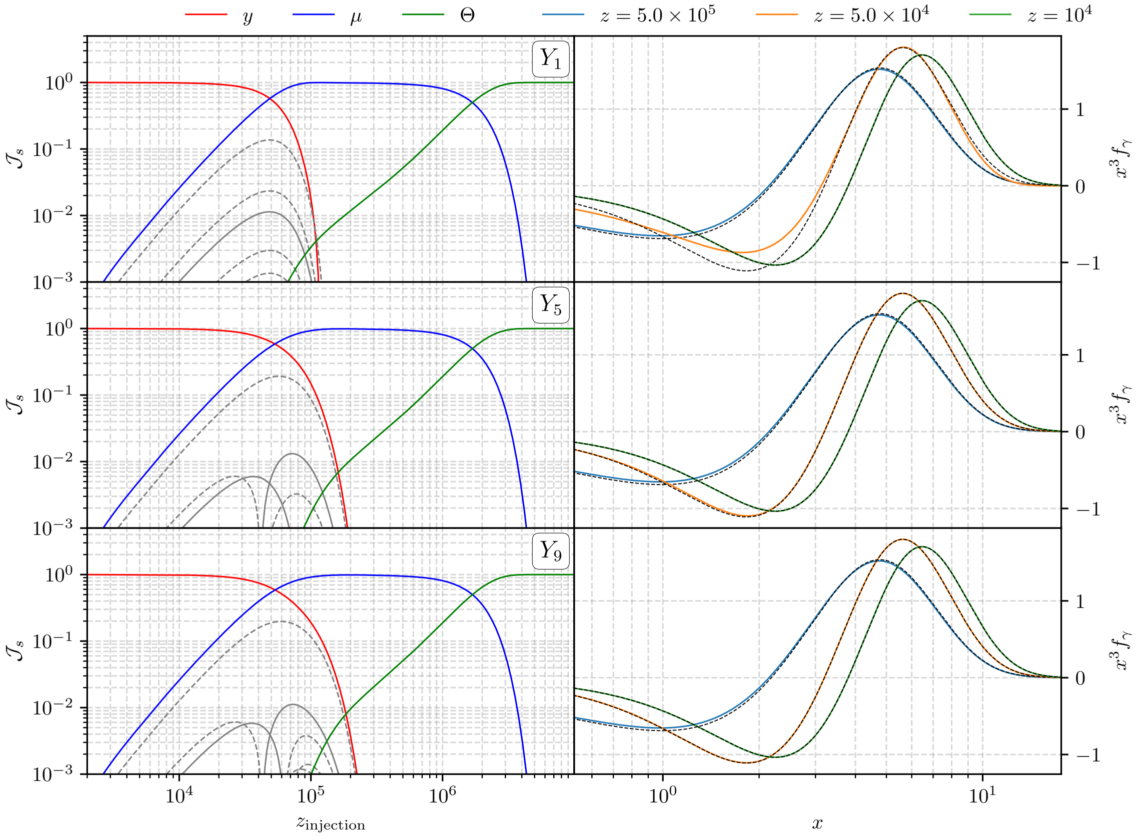


Figure 10: As for Fig. 4, but now with results cast into the “scattering” basis. Notice that in contrast to Fig. 7 there is no production of a negative temperature shift, since here we enforce a strict number conservation of the residual modes, meaning no distortion shape can project onto G in the change of basis. Again the spectra show no change compared to the computation basis.

density constraint. In the Y_{15} representation, one has $\Theta \approx 2.6 \times 10^{-12}$, $y \approx 6.1 \times 10^{-6}$ and $\mu \approx 3.0 \times 10^{-8}$ as before. Taking the full spectrum and imposing the photon number constraint to obtain the amplitude of G and then fitting for y and μ, we obtain $\Theta \approx 2.6 \times 10^{-12}$, $y \approx 1.6 \times 10^{-6}$ and $\mu \approx 7.7 \times 10^{-6}$. Just like before, we see a significant change in the values for y and μ, but this time no change to Θ. The total energy carried by G, Y and M is $\Delta\rho_\gamma/\rho_\gamma \simeq 4y + \mu/1.4 \approx 1.2 \times 10^{-5}$, implying that the residual distortion contributes $\Delta\rho_\gamma/\rho_\gamma \simeq -0.2 \times 10^{-5}$. In contrast, for the Y_{15} representation we have $\Delta\rho_\gamma/\rho_\gamma \approx 2.5 \times 10^{-5}$ stored in the G, Y and M components, implying that about $\Delta\rho_\gamma/\rho_\gamma \approx -1.5 \times 10^{-5}$ is in the $Y_{k>0}$ terms, which is *no* small total correction. If we compare all this to the lowest order computation using only Θ, y and μ in the ODE (i.e., a Y_0-representation) we obtain $\Theta \approx 1.1 \times 10^{-9}$, $y \approx 1.6 \times 10^{-6}$ and $\mu \approx 5.1 \times 10^{-6}$. This demonstrates that the μ and y decomposition is well captured by the alternative distortion eigenmodes.

However, in the outlined alternative procedure an observer has to evaluate the number integral $\propto \int x^2 \Delta n_x \,\mathrm{d}x$ of the photon field, which in experimental settings has several challenges. First, unless the distortion is measured in a sufficiently wide range of frequencies, this number integral would not evaluate accurately. Specifically, over a finite frequency domain even $\int x^2 Y \,\mathrm{d}x$ and $\int x^2 M \,\mathrm{d}x$ are no longer guaranteed to vanish, thereby breaking the ‘photon number orthogonality’. Second, to carry out the integral, the frequency sampling has to be fine, which again usually runs into observational

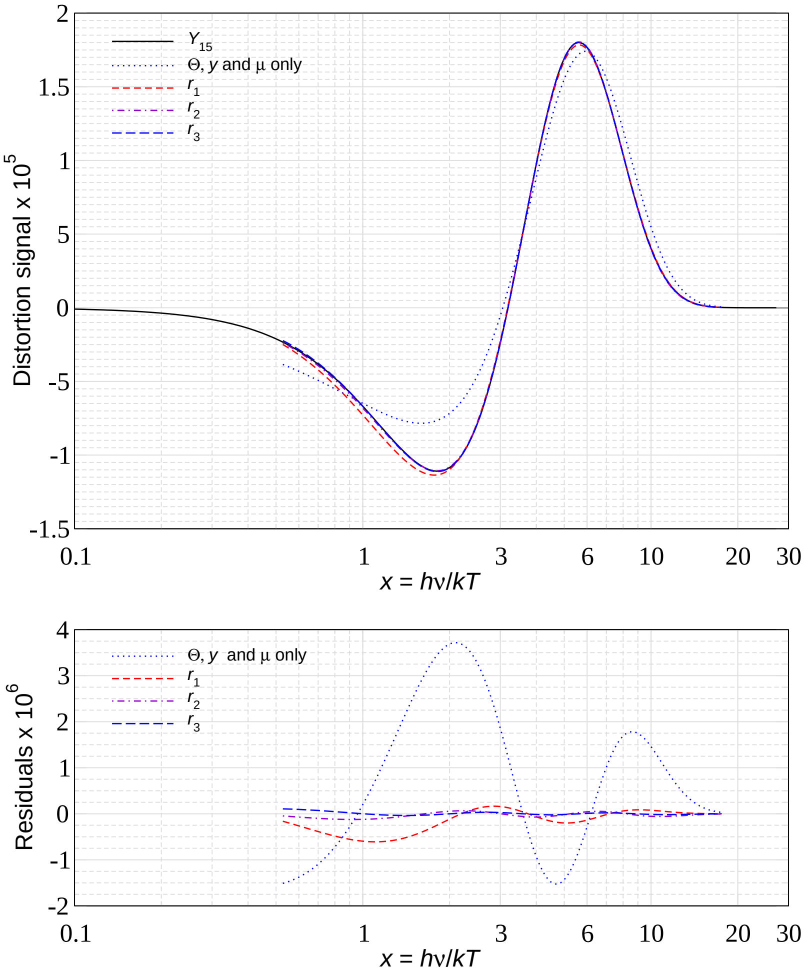


Figure 11: Same as Fig. 8 but using a photon number constraint to obtain the value for Θ.

difficulties. Third, the estimation of errors will be non-standard since the observable is based on weighted sums of fluxes. Therefore, this approach is not expected to be realized in actual observations. Nevertheless, for theoretical calculations, we can use it for illustration when the focus is on the energetics of the problem. We will therefore refer to this alternative basis as *scattering basis*, given that it is constructed to focus on the spectral shapes that are introduced purely by Compton scattering terms, which conserve photon number. We further discuss the benefits and differences of changing the basis in paper III.

4 Discussion and conclusions

In this work we obtained an approximate ODE treatment for the thermalisation Green's function for heating processes, which captures most aspects of the full `CosmoTherm` calculation using an extended spectral basis to describe the residual distortion evolution (Sect. 2). Instead of the expensive 'top-hat' frequency binning we use a spectral basis that is derived from boosts of the *y*-distortion spectrum. This reduces the computational burden by a factor of $\simeq 10^3$, thereby providing one of the main steps towards formulating a generalised photon Boltzmann hierarchy, that will allow us to compute the evolution of distortion anisotropies at first order in perturbation theory (see papers II and III). We also clarify how the computational spectral basis can be compressed into fewer distortion shapes that can be distinguished with a given experimental configuration, introducing the *observation* and *scattering* basis (Sect. 3).

The new ODE representation of the thermalisation Green's function given here is not perfect, because not all the spectral shapes that are excited by heating processes can be spanned by the basis functions we choose (see Fig. 5). However, we have demonstrated that this is not a severe problem for the average distortion evolution, which specifically relies on conversion of $Y(x)$ [as the main distortion source] to $M(x)$ and $G(x)$. The representation of the full Green's function could probably be improved by studying the eigenfunctions of the Kompaneets and boost operators more carefully. In addition, weighting schemes and a modified truncation of the distortion basis could likely improve the performance. One could also refine the treatment of photon emission processes, including the effect of the new distortion shapes. This is expected to modify the thermalisation efficiency, a problem that may be solved perturbatively. Nevertheless, the novel ODE representation of the average thermalisation Green's function is sufficiently accurate for approximate applications to SD anisotropies caused by energy release, as we show in papers II and III.

It is very important to highlight that not all spectral distortion problems can be treated this way. First, if the distortions are large, a full non-linear thermalisation problem is encountered, which require several extensions [11, 78]. However, given the distortion limits by *COBE/FIRAS*, this situation is usually not relevant to the description of primordial distortions. Second, the Kompaneets equation itself does not describe the general scattering problem in an isotropic medium. At high energies, the full Compton kernel has to be applied, which can cause noticeable departures to the evolution [e.g., 63]. Similarly, non-thermal distortions can in principle be excited by energy injection from particle cascades [78–80], which cannot be treated with the presented formalism. And finally, distortions created by direct photon injection can have a much richer spectral structure [73, 81] that cannot be captured here. We leave these cases for future studies, noting that at least for linear problems extended spectral bases should provide a means forward.

Overall, this paper is the first in a series of works discussing the evolution of SD anisotropies generated by various physical mechanisms and how these might be constrained with future CMB spectrometers and imagers. The results from these works should open the path for more realistic SD anisotropy forecasts over a wide range of physics which previously were not possible. This will hopefully spur additional activity on CMB spectral distortions, uniting the efforts of CMB imaging and spectrometer approaches for probing the early Universe.

Acknowledgments

We thank the referees of this paper and of the paper II in our series, for patiently digesting our work and whose comments helped us to clarify many important points. We also thank Eiichiro Komatsu, Aditya Rotti and Rashid Sunyaev for stimulating discussion at early stages. We furthermore thank Nicola Bartolo, Richard Battye, Daniele Bertacca, William Coulton, Bryce Cyr, Colin Hill, Antony Lewis, Rishi Khatri, Sabino Matarrese, Atsuhisa Ota, Enrico Pajer, Cyril Pitrou and Nils Schöneberg for comments on the manuscript. This work was supported by the ERC Consolidator Grant *CMBSPEC* (No. 725456). TK was also supported by STFC grant ST/T506291/1. JC was furthermore supported by the Royal Society as a Royal Society University Research Fellow at the University of Manchester, UK (No. URF/R/191023). AR acknowledges support by the project "Combining Cosmic Microwave Background and Large Scale Structure data: an Integrated Approach for Addressing Fundamental Questions in Cosmology", funded by the MIUR Progetti di Ricerca di Rilevante Interesse Nazionale (PRIN) Bando 2017 - grant 2017YJYZAH.

References

[1] R. A. Sunyaev and Y. B. Zeldovich, *The interaction of matter and radiation in the hot model of the Universe, II*, *Astrophysics and Space Science* **7** (Apr., 1970) 20–30.

[2] R. A. Daly, *Spectral distortions of the microwave background radiation resulting from the damping of pressure waves*, *The Astrophysical Journal* **371** (Apr., 1991) 14–28.

[3] W. Hu, D. Scott, and J. Silk, *Power spectrum constraints from spectral distortions in the cosmic microwave background*, *The Astrophysical Journal Letters* **430** (July, 1994) L5–L8, [`astro-ph/9`].

[4] J. Chluba, R. Khatri, and R. A. Sunyaev, *CMB at 2 x 2 order: the dissipation of primordial acoustic waves and the observable part of the associated energy release*, *Monthly Notices of the Royal Astronomical Society* **425** (Sept., 2012) 1129–1169, [`arXiv:1202.0057`].

[5] J. Delabrouille, M. H. Abitbol, N. Aghanim, et al., *Microwave spectro-polarimetry of matter and radiation across space and time*, *Experimental Astronomy* **51** (June, 2021) 1471–1514.

[6] J. Chluba, M. H. Abitbol, N. Aghanim, et al., *New horizons in cosmology with spectral distortions of the cosmic microwave background*, *Experimental Astronomy* **51** (June, 2021) 1515–1554, [`arXiv:1909.01593`].

[7] J. Chluba and R. A. Sunyaev, *The evolution of CMB spectral distortions in the early Universe*, *Monthly Notices of the Royal Astronomical Society* **419** (Jan., 2012) 1294–1314, [`arXiv:1109.6552`].

[8] R. A. Sunyaev and R. Khatri, *Unavoidable CMB Spectral Features and Blackbody Photosphere of Our Universe*, *IJMPD* **22** (June, 2013) 30014, [`arXiv:1302.6553`].

[9] H. Tashiro, *CMB spectral distortions and energy release in the early universe*, *Prog. of Theo. and Exp. Physics* **2014** (June, 2014) 060000.

[10] M. Lucca, N. Schöneberg, D. C. Hooper, J. Lesgourgues, and J. Chluba, *The synergy between CMB spectral distortions and anisotropies*, *Journal of Cosmology and Astroparticle Physics* **2020** (Feb., 2020) 026, [`arXiv:1910.04619`].

[11] J. Chluba, A. Ravenni, and S. K. Acharya, *Thermalization of large energy release in the early Universe*, *Monthly Notices of the Royal Astronomical Society* **498** (Oct., 2020) 959–980, [`arXiv:2005.11325`].

[12] A. Kogut, D. J. Fixsen, D. T. Chuss, et al., *The Primordial Inflation Explorer (PIXIE): a nulling polarimeter for cosmic microwave background observations*, *Journal of Cosmology and Astroparticle Physics* **7** (July, 2011) 25–+, [`arXiv:1105.2044`].

[13] P. André, C. Baccigalupi, A. Banday, et al., *PRISM (Polarized Radiation Imaging and Spectroscopy Mission): an extended white paper*, *Journal of Cosmology and Astroparticle Physics* **2** (Feb., 2014) 6, [`arXiv:1310.1554`].

[14] A. Kogut, J. Chluba, D. J. Fixsen, S. Meyer, and D. Spergel, *The Primordial Inflation Explorer (PIXIE)*, in *SPIE Conference Series*, vol. 9904 of *Proc.SPIE*, p. 99040W, July, 2016.

[15] A. Kogut, M. H. Abitbol, J. Chluba, et al., *CMB Spectral Distortions: Status and Prospects*, in *Bulletin of the American Astronomical Society*, vol. 51, p. 113, Sept., 2019. `arXiv:1907.13195`.

[16] B. Maffei, M. H. Abitbol, N. Aghanim, et al., *BISOU: a balloon project to measure the CMB spectral distortions*, *arXiv e-prints* (Oct., 2021) arXiv:2111.00246, [`arXiv:2111.00246`].

[17] S. Masi, E. Battistelli, P. de Bernardis, et al., *The COSmic Monopole Observer (COSMO)*, *arXiv e-prints* (Oct., 2021) arXiv:2110.12254, [`arXiv:2110.12254`].

[18] J. Chluba, A. Ravenni, and T. Kite, *Spectro-spatial evolution of the CMB II: generalised Boltzmann hierarchy*, *arXiv e-prints* (Oct., 2022) arXiv:2210.15308, [`arXiv:2210.15308`].

[19] T. Kite, A. Ravenni, and J. Chluba, *Spectro-spatial evolution of the CMB III: transfer functions, power spectra and Fisher forecasts*, *arXiv e-prints* (Dec., 2022) arXiv:2212.02817, [`arXiv:2212.02817`].

[20] S. K. Acharya and J. Chluba, *CMB spectral distortions from continuous large energy release*, *arXiv e-prints* (Dec., 2021) arXiv:2112.06699, [arXiv:2112.06699].

[21] J. Lesgourgues, *The Cosmic Linear Anisotropy Solving System (CLASS) I: Overview*, *ArXiv:1104.2932* (Apr., 2011) [arXiv:1104.2932].

[22] A. Lewis, A. Challinor, and A. Lasenby, *Efficient Computation of Cosmic Microwave Background Anisotropies in Closed Friedmann-Robertson-Walker Models*, *The Astrophysical Journal* **538** (Aug., 2000) 473–476, [astro-ph/9911177].

[23] Y. B. Zeldovich and R. A. Sunyaev, *The Interaction of Matter and Radiation in a Hot-Model Universe*, *Astrophysics and Space Science* **4** (July, 1969) 301–316.

[24] R. A. Sunyaev and I. B. Zeldovich, *The velocity of clusters of galaxies relative to the microwave background - The possibility of its measurement*, *Monthly Notices of the Royal Astronomical Society* **190** (Feb., 1980) 413–420.

[25] E. Komatsu and T. Kitayama, *Sunyaev-Zeldovich Fluctuations from Spatial Correlations between Clusters of Galaxies*, *The Astrophysical Journal Letters* **526** (Nov., 1999) L1–L4, [astro-ph/9908087].

[26] A. Refregier, E. Komatsu, D. N. Spergel, and U.-L. Pen, *Power spectrum of the Sunyaev-Zel'dovich effect*, *Physical Review D* **61** (June, 2000) 123001, [astro-ph/9].

[27] E. Komatsu and U. Seljak, *The Sunyaev-Zel'dovich angular power spectrum as a probe of cosmological parameters*, *Monthly Notices of the Royal Astronomical Society* **336** (Nov., 2002) 1256–1270, [astro-ph/0].

[28] J. C. Hill and E. Pajer, *Cosmology from the thermal Sunyaev-Zel'dovich power spectrum: Primordial non-Gaussianity and massive neutrinos*, *Physical Review D* **88** (Sept., 2013) 063526, [arXiv:1303.4726].

[29] B. Bolliet, B. Comis, E. Komatsu, and J. F. Macías-Pérez, *Dark Energy from the Thermal Sunyaev Zeldovich Power Spectrum*, *Mon. Not. Roy. Astron. Soc.* **477** (2018), no. 4 4957–4967, [arXiv:1712.00788].

[30] J. E. Carlstrom, G. P. Holder, and E. D. Reese, *Cosmology with the Sunyaev-Zel'dovich Effect*, *ARAA* **40** (2002) 643–680, [astro-ph/0].

[31] T. Mroczkowski, D. Nagai, K. Basu, et al., *Astrophysics with the Spatially and Spectrally Resolved Sunyaev-Zeldovich Effects. A Millimetre/Submillimetre Probe of the Warm and Hot Universe*, *Space Science Reviews* **215** (Feb., 2019) 17, [arXiv:1811.02310].

[32] A. Stebbins, *CMB Spectral Distortions from the Scattering of Temperature Anisotropies*, *astro-ph/0703541* (Mar., 2007) [astro-ph/0703541].

[33] C. Pitrou, F. Bernardeau, and J.-P. Uzan, *The y-sky: diffuse spectral distortions of the cosmic microwave background*, *Journal of Cosmology and Astroparticle Physics* **7** (July, 2010) 19, [arXiv:0912.3655].

[34] A. Ota, *CMB spectral distortions as solutions to the Boltzmann equations*, *Journal of Cosmology and Astroparticle Physics* **2017** (Jan., 2017) 037, [arXiv:1611.08058].

[35] E. Pajer and M. Zaldarriaga, *New Window on Primordial Non-Gaussianity*, *Physical Review Letters* **109** (July, 2012) 021302, [arXiv:1201.5375].

[36] J. Ganc and E. Komatsu, *Scale-dependent bias of galaxies and μ-type distortion of the cosmic microwave background spectrum from single-field inflation with a modified initial state*, *Physical Review D* **86** (July, 2012) 023518, [arXiv:1204.4241].

[37] M. Biagetti, H. Perrier, A. Riotto, and V. Desjacques, *Testing the running of non-Gaussianity through the CMB μ-distortion and the halo bias*, *Physical Review D* **87** (Mar., 2013) 063521, [arXiv:1301.2771].

[38] R. Emami, E. Dimastrogiovanni, J. Chluba, and M. Kamionkowski, *Probing the scale dependence of non-Gaussianity with spectral distortions of the cosmic microwave background*, *Physical Review D* **91** (June, 2015) 123531, [`arXiv:1504.00675`].

[39] A. Ota, *Cosmological constraints from μ E cross correlations*, *Physical Review D* **94** (Nov., 2016) 103520, [`arXiv:1607.00212`].

[40] J. Chluba, E. Dimastrogiovanni, M. A. Amin, and M. Kamionkowski, *Evolution of CMB spectral distortion anisotropies and tests of primordial non-Gaussianity*, *Monthly Notices of the Royal Astronomical Society* **466** (Apr., 2017) 2390–2401, [`arXiv:1610.08711`].

[41] A. Ravenni, M. Liguori, N. Bartolo, and M. Shiraishi, *Primordial non-Gaussianity with μ-type and y-type spectral distortions: exploiting Cosmic Microwave Background polarization and dealing with secondary sources*, *Journal of Cosmology and Astroparticle Physics* **2017** (Sept., 2017) 042, [`arXiv:1707.04759`].

[42] T. Haga, K. Inomata, A. Ota, and A. Ravenni, *Exploring compensated isocurvature perturbations with CMB spectral distortion anisotropies*, *Journal of Cosmology and Astroparticle Physics* **2018** (Aug., 2018) 036, [`arXiv:1805.08773`].

[43] M. Remazeilles, A. Ravenni, and J. Chluba, *Leverage on small-scale primordial non-Gaussianity through cross-correlations between CMB E-mode and μ-distortion anisotropies*, *Monthly Notices of the Royal Astronomical Society* **512** (May, 2022) 455–470, [`arXiv:2110.14664`].

[44] A. Rotti, A. Ravenni, and J. Chluba, *Non-Gaussianity constraints from Planck spectral distortion cross-correlations*, *arXiv e-prints* (May, 2022) arXiv:2205.15971, [`arXiv:2205.15971`].

[45] F. Bianchini and G. Fabbian, *CMB spectral distortions revisited: A new take on <math display="inline"><mi>μ</mi></math> distortions and primordial non-Gaussianities from FIRAS data*, *Phys. Rev. D* **106** (2022), no. 6 063527, [`arXiv:2206.02762`].

[46] E. Pajer and M. Zaldarriaga, *A hydrodynamical approach to CMB μ-distortion from primordial perturbations*, *Journal of Cosmology and Astroparticle Physics* **2** (Feb., 2013) 36, [`arXiv:1206.4479`].

[47] J. C. Mather, E. S. Cheng, D. A. Cottingham, et al., *Measurement of the cosmic microwave background spectrum by the COBE FIRAS instrument*, *The Astrophysical Journal* **420** (Jan., 1994) 439–444.

[48] D. J. Fixsen, E. S. Cheng, J. M. Gales, et al., *The Cosmic Microwave Background Spectrum from the Full COBE FIRAS Data Set*, *The Astrophysical Journal* **473** (Dec., 1996) 576–+, [`astro-ph/9`].

[49] D. Zegeye, K. Inomata, and W. Hu, *Spectral distortion anisotropy from inflation for primordial black holes*, *Physical Review D* **105** (May, 2022) 103535, [`arXiv:2112.05190`].

[50] J. Chluba and D. Jeong, *Teasing bits of information out of the CMB energy spectrum*, *Monthly Notices of the Royal Astronomical Society* **438** (Mar., 2014) 2065–2082, [`arXiv:1306.5751`].

[51] D. J. Fixsen, *The Temperature of the Cosmic Microwave Background*, *The Astrophysical Journal* **707** (Dec., 2009) 916–920, [`arXiv:0911.1955`].

[52] J. Chluba, *Future Steps in Cosmology using Spectral Distortions of the Cosmic Microwave Background*, *arXiv e-prints* (June, 2018) arXiv:1806.02915, [`arXiv:1806.02915`].

[53] J. Chluba, *Green's function of the cosmological thermalization problem*, *Monthly Notices of the Royal Astronomical Society* **434** (Sept., 2013) 352–357, [`arXiv:1304.6120`].

[54] J. Chluba, *Which spectral distortions does ΛCDM actually predict?*, *Monthly Notices of the Royal Astronomical Society* **460** (July, 2016) 227–239, [`arXiv:1603.02496`].

[55] R. Khatri and R. A. Sunyaev, *Beyond y and μ: the shape of the CMB spectral distortions in the intermediate epoch, 1.5 x 10^4lesssimzlesssim2 x 10^5*, *Journal of Cosmology and Astroparticle Physics* **9** (Sept., 2012) 16, [`arXiv:1207.6654`].

[56] L. Danese and G. de Zotti, *Double Compton process and the spectrum of the microwave background*, *Astronomy & Astrophysics* **107** (Mar., 1982) 39–42.

[57] C. Burigana, L. Danese, and G. de Zotti, *Formation and evolution of early distortions of the microwave background spectrum - A numerical study*, *Astronomy & Astrophysics* **246** (June, 1991) 49–58.

[58] W. Hu and J. Silk, *Thermalization and spectral distortions of the cosmic background radiation*, *Physical Review D* **48** (July, 1993) 485–502.

[59] S. Y. Sazonov and R. A. Sunyaev, *Gas Heating Inside Radio Sources to Mildly Relativistic Temperatures via Induced Compton Scattering*, *Astronomy Letters* **27** (Aug., 2001) 481–492, [astro-ph/0].

[60] A. Sarkar, J. Chluba, and E. Lee, *Dissecting the Compton scattering kernel I: Isotropic media*, *Monthly Notices of the Royal Astronomical Society* **490** (Dec., 2019) 3705–3726, [arXiv:1905.00868].

[61] J. Chluba, *Refined approximations for the distortion visibility function and μ-type spectral distortions*, *Monthly Notices of the Royal Astronomical Society* **440** (Apr., 2014) 2544–2563, [arXiv:1312.6030].

[62] J. Chluba and R. A. Sunyaev, *Evolution of low-frequency features in the CMB spectrum due to stimulated Compton scattering and Doppler broadening*, *Astronomy & Astrophysics* **488** (Sept., 2008) 861–865, [arXiv:0804.1017].

[63] S. K. Acharya, J. Chluba, and A. Sarkar, *Comparison of numerical methods for computing the repeated Compton scattering of photons in isotropic media*, *Monthly Notices of the Royal Astronomical Society* (Aug., 2021) [arXiv:2105.04496].

[64] C. Pitrou and A. Stebbins, *Parameterization of temperature and spectral distortions in future CMB experiments*, *arXiv e-prints* (Feb., 2014) arXiv:1402.0968, [arXiv:1402.0968].

[65] N. Itoh, Y. Kohyama, and S. Nozawa, *Relativistic Corrections to the Sunyaev-Zeldovich Effect for Clusters of Galaxies*, *The Astrophysical Journal* **502** (July, 1998) 7–+, [astro-ph/9].

[66] S. Y. Sazonov and R. A. Sunyaev, *Cosmic Microwave Background Radiation in the Direction of a Moving Cluster of Galaxies with Hot Gas: Relativistic Corrections*, *The Astrophysical Journal* **508** (Nov., 1998) 1–5.

[67] A. Challinor and A. Lasenby, *Relativistic Corrections to the Sunyaev-Zeldovich Effect*, *The Astrophysical Journal* **499** (May, 1998) 1–+, [astro-ph/9].

[68] J. Chluba, D. Nagai, S. Sazonov, and K. Nelson, *A fast and accurate method for computing the Sunyaev-Zel'dovich signal of hot galaxy clusters*, *Monthly Notices of the Royal Astronomical Society* **426** (Oct., 2012) 510–530, [arXiv:1205.5778].

[69] C. Pitrou, *The radiative transfer at second order: a full treatment of the Boltzmann equation with polarization*, *Classical and Quantum Gravity* **26** (Mar., 2009) 065006, [arXiv:0809.3036].

[70] A. Ota, T. Takahashi, H. Tashiro, and M. Yamaguchi, *CMB μ distortion from primordial gravitational waves*, *Journal of Cosmology and Astroparticle Physics* **10** (Oct., 2014) 29, [arXiv:1406.0451].

[71] A. Ravenni and J. Chluba, *The double Compton process in astrophysical plasmas*, *Journal of Cosmology and Astroparticle Physics* **2020** (Oct., 2020) 025, [arXiv:2005.06941].

[72] J. Chluba, A. Ravenni, and B. Bolliet, *Improved calculations of electron-ion bremsstrahlung Gaunt factors for astrophysical applications*, *Monthly Notices of the Royal Astronomical Society* **492** (Feb., 2020) 177–194, [arXiv:1911.08861].

[73] J. Chluba, *Green's function of the cosmological thermalization problem - II. Effect of photon injection and constraints*, *Monthly Notices of the Royal Astronomical Society* **454** (Dec., 2015) 4182–4196, [arXiv:1506.06582].

[74] J. Chluba, *Distinguishing different scenarios of early energy release with spectral distortions of the cosmic microwave background*, *Monthly Notices of the Royal Astronomical Society* **436** (Dec., 2013) 2232–2243, [arXiv:1304.6121].

[75] E. Hivon, K. M. Górski, C. B. Netterfield, et al., *MASTER of the Cosmic Microwave Background Anisotropy Power Spectrum: A Fast Method for Statistical Analysis of Large and Complex Cosmic*

Microwave Background Data Sets, *The Astrophysical Journal* **567** (Mar., 2002) 2–17, [astro-ph/0105302].

[76] M. H. Abitbol, J. Chluba, J. C. Hill, and B. R. Johnson, *Prospects for Measuring Cosmic Microwave Background Spectral Distortions in the Presence of Foregrounds*, *Monthly Notices of the Royal Astronomical Society* (May, 2017) [arXiv:1705.01534].

[77] A. Rotti and J. Chluba, *Combining ILC and moment expansion techniques for extracting average-sky signals and CMB anisotropies*, *Monthly Notices of the Royal Astronomical Society* **500** (Jan., 2021) 976–985, [arXiv:2006.02458].

[78] S. K. Acharya and R. Khatri, *Rich structure of nonthermal relativistic CMB spectral distortions from high energy particle cascades at redshifts z <2 x 10^5*, *Physical Review D* **99** (Feb., 2019) 043520, [arXiv:1808.02897].

[79] T. A. Enßlin and C. R. Kaiser, *Comptonization of the cosmic microwave background by relativistic plasma*, *Astronomy & Astrophysics* **360** (Aug., 2000) 417–430, [astro-ph/0].

[80] T. R. Slatyer, *Indirect dark matter signatures in the cosmic dark ages. II. Ionization, heating, and photon production from arbitrary energy injections*, *Physical Review D* **93** (Jan, 2016) 023521, [arXiv:1506.03812].

[81] B. Bolliet, J. Chluba, and R. Battye, *Spectral distortion constraints on photon injection from low-mass decaying particles*, *arXiv e-prints* (Dec., 2020) arXiv:2012.07292, [arXiv:2012.07292].

[82] J. Chluba, D. Nagai, S. Sazonov, and K. Nelson, *A fast and accurate method for computing the Sunyaev-Zel'dovich signal of hot galaxy clusters*, *Monthly Notices of the Royal Astronomical Society* **426** (Oct., 2012) 510–530, [arXiv:1205.5778].

A Useful operator properties

We can somewhat reduce the complexity of the above calculations by studying the properties of – and relationships between – the operators $\hat{\mathcal{D}}_x$, $\hat{\mathcal{D}}_x^*$ and $\hat{O}_x$. This will furthermore illustrate the suitability of the expanded basis $\{Y_k(x)\}$.

We first note that two of the main operators commute with one another:

$$[\hat{\mathcal{D}}_x, \hat{O}_x] = 0, \tag{A.1}$$

thus implying the potential existence of a shared eigenbasis. A logical step is to express the *larger* of the operators in terms of the other, revealing the identities

$$\hat{\mathcal{D}}_x = \hat{O}_x(\hat{O}_x - 3) = (\hat{O}_x - 3)\hat{O}_x = -3\hat{O}_x + (\hat{O}_x)^2 = 4x\partial_x + x^2\,\partial_x^2. \tag{A.2}$$

The final equality can be easily found with a single application of the chain rule. However, it hints towards a more generic recurrence relation, which yields the following combinatoric sum:

$$(\hat{O}_x)^k = (-1)^k \sum_{m=1}^{k} \begin{bmatrix} k \\ m \end{bmatrix} x^m \partial_x^m, \tag{A.3}$$

where the square brackets indicate *Stirling set numbers*, which counts partitions of an k-set into m nonempty subsets. This expansion of the boost operator reveals that the reverse operation is non-trivial – very specific weighted sums of $x^m\partial_x^m$ terms are needed to make a boost operator with some power. Because of this, it is useful to be able to compose these expanded $\hat{O}_x$ terms directly:

$$x^a\,\partial_x^a\,x^b\,\partial_x^b = \sum_{k=0}^{a} \frac{b!}{(b-a+k)!} \binom{a}{k} x^{b+k}\,\partial_x^{b+k}. \tag{A.4}$$

As noted in [82], $x^k\partial_x^k n_{\rm bb}$ has a recursion relation allowing for another combinatoric analytic solution

$$x^k\partial_x^k n_{\rm bb} = \frac{(-x)^k \mathrm{e}^{-x}}{(1-\mathrm{e}^{-x})^{k+1}} \sum_{m=0}^{k-1} \left\langle {k \atop m} \right\rangle \mathrm{e}^{-mx} \qquad (k>0), \tag{A.5}$$

where the angle brackets denote *Eulerian numbers*, defined as the number of permutations of the numbers 1 to m in which exactly k elements are greater than the previous element. This expression has very good convergence properties when summed starting from the highest power of e^{-mx}.

Defining $\mathcal{H}_k(x) = (-x)^k \mathrm{e}^{-x}/(1-\mathrm{e}^{-x})^{k+1}$, we are now in a position to write general expressions using the above formulae:

$$\hat{\mathcal{D}}_x^N = \sum_{k=0}^{N} \binom{N}{k} 3^k \sum_{m=1}^{2N-k} \begin{bmatrix} 2N-k \\ m \end{bmatrix} x^m\,\partial_x^m, \tag{A.6}$$

$$(\hat{\mathcal{O}}_x)^N n_{\rm bb} = (-1)^N \sum_{k=1}^{N} \begin{bmatrix} N \\ k \end{bmatrix} \mathcal{H}_k(x) \sum_{m=0}^{k-1} \left\langle {k \atop m} \right\rangle \mathrm{e}^{-mx} \qquad (N>0), \tag{A.7}$$

$$\hat{\mathcal{D}}_x^N (\hat{\mathcal{O}}_x)^M n_{\rm bb} = (-1)^M \sum_{k=0}^{N} \binom{N}{k} 3^k \sum_{\ell=1}^{2N+M-k} \begin{bmatrix} 2N+M-k \\ \ell \end{bmatrix} \mathcal{H}_\ell \sum_{m=0}^{\ell-1} \begin{bmatrix} l \\ m \end{bmatrix} \mathrm{e}^{-mx}. \tag{A.8}$$

We can generate the basis functions Y_N from the above expressions noticing that according to our convention $Y_N = (\hat{\mathcal{O}}_x/4)^N Y = \hat{\mathcal{D}}_x(\hat{\mathcal{O}}_x/4)^N n_{\rm bb}$:

$$Y_N = (-1/4)^N \sum_{k=1}^{N+2} \left(3\begin{bmatrix} N+1 \\ k \end{bmatrix} + \begin{bmatrix} N+2 \\ k \end{bmatrix} \right) \mathcal{H}_k \sum_{m=0}^{k-1} \left\langle {k \atop m} \right\rangle \mathrm{e}^{-mx}. \tag{A.9}$$

Note that we have used $\begin{bmatrix} N+1 \\ N \end{bmatrix} = 0$ to simplify the above expression, bringing two different powers of derivatives under a single summation sign. Below we provide a few examples:

$$Y_1(x) = \frac{\mathrm{e}^{-x}\left(\mathrm{e}^{-2x}+4\mathrm{e}^{-x}+1\right)x^3}{4\,(1-\mathrm{e}^{-x})^4} - \frac{3\mathrm{e}^{-x}\,(\mathrm{e}^{-x}+1)\,x^2}{2\,(1-\mathrm{e}^{-x})^3} + \frac{\mathrm{e}^{-x}x}{(1-\mathrm{e}^{-x})^2}, \tag{A.10}$$

$$\begin{aligned} Y_2(x) = & \frac{\mathrm{e}^{-x}\left(\mathrm{e}^{-3x}+11\mathrm{e}^{-2x}+11\mathrm{e}^{-x}+1\right)x^4}{16\,(1-\mathrm{e}^{-x})^5} - \frac{9\mathrm{e}^{-x}\left(\mathrm{e}^{-2x}+4\mathrm{e}^{-x}+1\right)x^3}{16\,(1-\mathrm{e}^{-x})^4} \\ & + \frac{\mathrm{e}^{-x}\,(\mathrm{e}^{-x}+1)\,x^2}{(1-\mathrm{e}^{-x})^3} - \frac{\mathrm{e}^{-x}x}{4\,(1-\mathrm{e}^{-x})^2}, \end{aligned} \tag{A.11}$$

$$\begin{aligned} Y_3(x) = & \frac{\mathrm{e}^{-x}\left(\mathrm{e}^{-4x}+26\mathrm{e}^{-3x}+66\mathrm{e}^{-2x}+26\mathrm{e}^{-x}+1\right)x^5}{64\,(1-\mathrm{e}^{-x})^6} - \frac{13\mathrm{e}^{-x}\left(\mathrm{e}^{-3x}+11\mathrm{e}^{-2x}+11\mathrm{e}^{-x}+1\right)x^4}{64\,(1-\mathrm{e}^{-x})^5} \\ & + \frac{43\mathrm{e}^{-x}\left(\mathrm{e}^{-2x}+4\mathrm{e}^{-x}+1\right)x^3}{64\,(1-\mathrm{e}^{-x})^4} - \frac{9\mathrm{e}^{-x}\,(\mathrm{e}^{-x}+1)\,x^2}{16\,(1-\mathrm{e}^{-x})^3} + \frac{\mathrm{e}^{-x}x}{16\,(1-\mathrm{e}^{-x})^2}. \end{aligned} \tag{A.12}$$

Despite the progress made above, the overall problem is not fully closed via combinatoric sums. The operator $\mathcal{D}_x^* \in \hat{\mathcal{K}}_x$ does not commute with the others. Instead we find

$$\left[\frac{1}{x}\hat{\mathcal{D}}_x^*, \hat{\mathcal{O}}_x\right] = 0, \qquad \left[\hat{\mathcal{D}}_x^*\frac{1}{x}, \hat{\mathcal{O}}_x\right] = 0, \qquad \left[\hat{\mathcal{D}}_x, \frac{1}{x}\hat{\mathcal{D}}_x^*\right] = 0, \qquad \left[\mathcal{D}_x, \hat{\mathcal{D}}_x^*\frac{1}{x}\right] = 0, \tag{A.13}$$

showing that no such shared basis will exist, and thus for now we resort to the approximate numerical projections discussed in the main text (see especially Fig. 2).

However, some more progress can be made my realising that $\hat{\mathcal{D}}_x^*$ always appears in conjunction with with the factor $A = (1 + 2n_{\rm bb})$. It can be shown that $A = \frac{1}{x}\left(4 + \frac{Y}{G}\right)$. This loose factor of $1/x$ combines nicely with the commutators noted above. Specifically we can then write

$$\hat{\mathcal{D}}_x^* A = -(\hat{O}_x - 3)\left(4 + \frac{Y}{G}\right). \tag{A.14}$$

Combining this with Eq.(A.2) we can write

$$\hat{\mathcal{K}}_x = (\hat{O}_x - 3)\left[\hat{O}_x - 4 - \frac{Y}{G}\right] = \hat{O}_x^2 - 7\hat{O}_x + 12 + 3\frac{Y}{G} - \hat{O}_x\frac{Y}{G}, \tag{A.15}$$

which essentially distils the *misbehaving* part of the Kompaneets operator to the previously named y-weight factor $w_y = Y/G$.

This expression of the Kompaneets operator makes it clearer to see how certain results arise algebraically. Consider for example that $\hat{\mathcal{K}}_x G = -Y$, and similarly $\hat{\mathcal{K}}_x M = -Y$, where the latter result follows from the former together with $\hat{\mathcal{K}}_x(G/x) = 0$. The spectral shape, Y_1, appears as intermediate step in these calculations, but ends up cancelling. These results may not be interesting in isolation, but they emphasise the fact that the cancellations only occur for simple shapes. Once you apply $\hat{\mathcal{K}}_x$ to a distortion shape like Y you naturally get Y_1 and Y_2 that do not analytically cancel.

B Alternative derivation of the ODE system

To obtain the ODE system for the evolution of the spectrum, we can also directly project the evolution equation. Making the Ansatz $\Delta n \approx \boldsymbol{B} \cdot \boldsymbol{y}$ (with definitions as in the main section for a given basis) and then inserting this into the evolution equation, Eq. (2.7), we have

$$\Theta' G(x) + \sum_{k=0}^{N} y_k' Y_k(x) + \mu' M(x) = \Theta_{\rm e}\, Y(x) - \Theta\, Y(x) + \sum_{k=0}^{N} y_k\, K_{Y_k}(x) - \mu\, \eta_M Y(x). \tag{B.1}$$

Here, $Y_0 \equiv Y$, $K_{Y_k} = \hat{\mathcal{K}}_x Y_k$ and we used the identities in Eq. (2.12). Since only $G(x)$ carries number we immediately obtain $\Theta' = 0$ by carrying out the number integral $\int x^2\,{\rm d}x$ over this equation. Since we know that $-\Theta\, Y(x)$ on the right hand side of Eq. (B.1) cancels the corresponding term in the Compton equilibrium temperature

$$\Theta_{\rm e} \approx \left(\frac{\int x^3 w_y(x)\, \boldsymbol{B}\, {\rm d}x}{4E_{n_{\rm bb}}}\right) \cdot \boldsymbol{y} \approx \Theta + \sum_{k=0}^{N} \eta_{Y_k}\, y_k + \eta_M\, \mu, \tag{B.2}$$

and because there also is no term $\propto G(x)$, we only have to worry about the reduced problem

$$\sum_{k=0}^{N} y_k' Y_k(x) + \mu' M(x) = (\Theta_{\rm e} - \Theta - \eta_M\, \mu)\, Y(x) + \sum_{k=0}^{N} y_k\, K_{Y_k}(x). \tag{B.3}$$

By performing the projections onto all function of the representation basis $\boldsymbol{R} = (Y, Y_1, \ldots, Y_k, M)^T$, we obtain the system

$$M_R\, \boldsymbol{y}' = (\Theta_{\rm e} - \Theta - \mu\, \eta_M)\, \boldsymbol{b}_Y + K\, \boldsymbol{y}. \tag{B.4}$$

where $\boldsymbol{y} = (y, y_1, \ldots, y_k, \mu)^T$ and $M_{R,ij} = \langle R_i | R_j \rangle$ is the full mixing matrix. We also have the source vector $b_{Y,i} = \langle R_i | R_0 \rangle = \langle R_i | Y \rangle$ and Kompaneets matrix $K_{ij} = \langle R_i | K_{Y_j} \rangle$.

As already explained in the main text, the system above will not yield a solution that correctly conserves energy (although it will become better and better the more Y_k are included). We therefore replace the last row in the matrices M_R and K and the last entry in $\boldsymbol{b}_Y$ with the corresponding energy equation (as shown in the main text). The modified system has the same form as Eq. (B.4), just with redefined matrices and vectors which we do not explicitly distinguish in the notation. The system can be solved for $\boldsymbol{y}'$ to obtain the evolution equation for $y, y_1, \ldots, y_N, \mu$ as

$$\boldsymbol{y}' = (\Theta_\mathrm{e} - \Theta - \mu\eta_M)\; M_R^{-1}\, \boldsymbol{b}_Y + M_R^{-1}\, K\, \boldsymbol{y}. \tag{B.5}$$

The rows of the matrix $M_R^{-1}\,K$ are composed of the representation vectors for the operators K_{Y_k}. Note that the matrix K is an $(N+2) \times (N+1)$ matrix, while ${M_R}^{-1}$ is an $(N+2) \times (N+2)$ matrix, such that $M_R^{-1}\,K$ also is an $(N+2) \times (N+1)$ matrix. In addition we have $M_R^{-1}\,\boldsymbol{b}_Y = \delta_{i0}$, which simply follows from the fact that $\tilde{\boldsymbol{b}}_Y$ is the first column vector of the matrix M_R. Since the matrix $M_R^{-1}\,K$ can be determined by independently solving for the representations of K_{Y_k} in terms of the representation basis $\boldsymbol{R}$, this means we have proven the equivalence with the approach used in the main text.

7

Paper IV: Spectro-spatial evolution of the CMB II: generalised Boltzmann hierarchy

This paper is the second in a series of three papers (see chapter 6 and chapter 8) which together aim to correctly model the frequency domain for the photon phase space distribution (see chapter 2) within the Boltzmann hierarchy (see chapter 3).

With the spectral basis proven to work for the homogeneous Universe, this paper explores the extended perturbation theory once the discrete frequency basis has been included (i.e. the calculations outlined in Sect. 3.2 but generalising from $f^{(1)} = \Theta^{(1)}\mathcal{G}$).

The main body of text builds the picture in a more direct and pedagogical way, including the construction of specific matrices and vectors for computation, before presenting a bottom line equation which is a direct counterpart to the usual photon hierarchy. The appendices explore different aspects of the perturbations to much greater depth, which all feed into the bottom line result in the main text.

Importantly we show that local distorted photon patches arise from a combination of an average background distorted photon spectrum and local inflationary perturbations. This crucially means that inspecting the distorted anisotropies in the CMB sky reveal details about the average spectrum, thus providing a completely novel approch to contraining average energy injection in the primordial plasma. While some aspects of the solutions are discussed, the results are left to the following paper.

Prepared for submission to JCAP

Spectro-spatial evolution of the CMB II: generalised Boltzmann hierarchy

Jens Chluba[a] , **Andrea Ravenni**[b,c] **and Thomas Kite**[a]

[a]Jodrell Bank Centre for Astrophysics, School of Physics and Astronomy, The University of Manchester, Oxford Road, Manchester, M13 9PL, U.K.

[b]Dipartimento di Fisica e Astronomia "Galileo Galilei", Università degli Studi di Padova, via Marzolo 8, I-35131, Padova, Italy.

[c]INFN, Sezione di Padova, via Marzolo 8, I-35131, Padova, Italy.

E-mail: Jens.Chluba@Manchester.ac.uk, Andrea.Ravenni@unipd.it,
Thomas.Kite@Manchester.ac.uk

Abstract. In this paper, we formulate a generalised photon Boltzmann hierarchy that allows us to model the evolution and creation of spectral distortion anisotropies caused by energy release in the early Universe. We directly build on our first paper in this series, extending the thermalisation Green's function treatment to the anisotropic case. We show that the problem can be described with the common Boltzmann hierarchy for the photon field extended by new spectral parameters – a step that reduces the complexity of the calculation by at least two orders of magnitude. Our formalism describes the effects of i) Doppler and potential driving, ii) spectral evolution by Compton scattering, iii) perturbed thermalisation and iv) anisotropic heating on the distortion anisotropies. We highlight some of the main physical properties of the equations and also outline the steps for computing CMB power spectra including distortion anisotropies. Limitations and extensions of the formulation are also briefly discussed. The novel Boltzmann hierarchy given here is the basis for a series of companion papers studying how distortion anisotropies evolve in the perturbed Universe and which physical processes could be constrained using future CMB imaging techniques.

Contents

1 Introduction 2

2 Extended Boltzmann hierarchy with primordial spectral distortions 2
- 2.1 Paper I and zeroth order problem 3
- 2.2 General statement of the problem in first order of perturbation theory 4
- 2.3 Effect of Doppler and potential terms in the Thomson limit 5
  - 2.3.1 Preparing for more general spectral evolution 5
  - 2.3.2 Changing the background temperature 7
  - 2.3.3 Instantaneous thermalisation 10
- 2.4 Effect of Compton scattering and photon production 11
  - 2.4.1 Simplistic inclusion of thermalisation terms 11
  - 2.4.2 More rigorous treatment 12
- 2.5 Final evolution equation 13
- 2.6 Effect of distortion on the other perturbation equations 13
- 2.7 Expression in Fourier space 14
- 2.8 Line of sight integration and power spectra 15
  - 2.8.1 Reduction to the experimental basis 16
- 2.9 Basic expectations for the evolution of distortion anisotropies 16

3 Discussion and conclusions 17

A Derivation of the photon collision term at first order 23
- A.1 Thomson terms 23
- A.2 Kompaneets terms 23
  - A.2.1 Kinematic corrections to the Kompaneets term 25
- A.3 Photon number changing processes 26

B Evolution in various limits 27
- B.1 Evolution equations in Thomson limit 27
- B.2 Changing the average temperature at second order in $\bar{\Theta}$ 27

C Physical approximations for the kinetic equation 28
- C.1 Compton Electron temperature and Compton energy exchange 28
- C.2 Electron temperature equation and effective photon heating rate 29
  - C.2.1 Adding extra Doppler terms to the system 31
  - C.2.2 Compton equilibrium spectrum 31
- C.3 Photon production in the μ-era 32
  - C.3.1 Zeroth order treatment of photon production 32
  - C.3.2 First order treatment of photon production 34

D Details of the derivation for the line-of-sight integral solution. 36

1 Introduction

The cosmic microwave background (CMB) has been a goldmine for furthering our understanding of the cosmos [1–5]. We are now entering a new phase, in which novel cosmological observables are moving to the focus of our efforts [6–9]. One of these observables is spectral distortions (SDs) of the CMB [10, 11]. In this work, we develop a novel Boltzmann formulation that can describe the evolution of SD *anisotropies* created by energy release in the early Universe. We base the treatment on our first paper in the series [12, henceforth referred to as paper I], which introduced a novel discretisation of the thermalisation Green's function to efficiently describe the evolution of average SDs. In paper I, we were limited to running one single thermalisation history at the time, effectively obtaining a solution for the average distortion intensity spectrum, $\Delta I_\nu(t)$, at a given time t and frequency ν. In simple words, we now deliver the tools to repeat the calculation along multiple lines of sight and including the effect of perturbations on the CMB signal evolution. This will open the way to studying SD physics using standard methods known from CMB imaging, without the need for absolute calibration, delivering new targets for experiments like *Litebird* [13], *PICO* [14] and CMB-S4 [6].

We refer the reader to paper I for more introduction and motivation to the topic. The steps taken here explain in detail how to extend the thermalisation Green's function treatment to the anisotropic case. We are in particular keen on including thermalisation effects to the evolution. This captures the full spectral evolution due to Compton scattering, double Compton (DC) and Bremsstrahlung (BR), which are so crucial in the formation of SDs at redshifts $z \gtrsim 10^4$ [15–19]. These effects can change the type of the distortion and gradually *rotate* y-type distortion sources into μ and ultimately thermalise the distortion completely. The rotation is with respect to the energy in various spectral components, which by construction is conserved across the spectral basis. A perturbative formulation that includes some of the Doppler boosting effects has been given in [20] and [21, 22]; however, this does not capture the effects of repeated Compton scattering at $z > 10^5$ and also does not describe the conversion of μ to T, thus having limited applicability, which we overcome here.

In Sect. 2, we start by recapping the main outcomes of paper I. We then move on to describing some of the distortion effects in the Thomson limit of the kinetic equation. This only captures the effect of Doppler and potential driving as well as free-streaming mixing on the SD signals across the sky, but provides valuable insight preparing for the general case. In Sect. 2.4, we add thermalisation terms to complete our treatment. Details of the derivation are giving in Appendix A to C, which we recommend for in depth reading. The tools for computing the CMB power spectra for various combinations of temperature and SD parameters are given in Sect. 2.8, and a discussion of the basic expectations is presented in Sect. 2.9. Numerical solutions and first forecasts will be presented in paper III [23]. In Sect. 3, we present our conclusions and also highlight some of the limitation of the formulation and further work that may become important.

2 Extended Boltzmann hierarchy with primordial spectral distortions

The goal of this section is the provide an approximate description of the spectro-spatial thermalisation problem at first order in perturbation theory. This greatly generalizes previous treatments of the problem allowing us to capture the main sources of distortions in the presence of perturbations. Unlike the usual split, here we have two small parameters to work with, one for the average energy release, $\epsilon_\gamma = \Delta\rho_\gamma/\rho_\gamma$, and one for the primordial curvature perturbations, ζ. We keep terms up to $O(\epsilon_\gamma \zeta)$, thereby allowing to capture all linear order (spatial) perturbation effects.

We will use the approach presented in paper I to model the evolution of the local monopole across the chosen spectral basis, while the spatial evolution is treated using the standard multipole decomposition which includes the effect of Thomson scattering and free-streaming. We will make direct use of the main results from paper I and refer the interested reader to that paper for clarification of the notation and details of the Green's function discretisation and choice of basis functions.

2.1 Paper I and zeroth order problem

In paper I, we essentially solved the zeroth order photon equation for the evolution of the average CMB spectrum. This problem can be described with the kinetic equation [see Sect. 4.1.1 of 24]:

$$\frac{1}{\dot{\tau}}\frac{\partial \Delta n^{(0)}}{\partial t} \approx \theta_z \Theta_{\rm eq}^{(0)}\, Y + \theta_z \hat{\mathcal{K}}_x\, \Delta n^{(0)} - \frac{\Lambda\,(1-{\rm e}^{-x})}{x^3} \Delta n^{(0)} + \frac{\Lambda}{x^2}\, n_{\rm bb}\, \Theta_{\rm eq}^{(0)} + \frac{\dot{Q}^{(0)}}{\dot{\tau}}\, Y. \tag{2.1}$$

for the evolution of the photon occupation number distortion $\Delta n^{(0)} \equiv \Delta n^{(0)}(t,x)$, defined with respect to the average blackbody, $n_{\rm bb}(x = h\nu/k_{\rm B}T_z) = 1/({\rm e}^x - 1)$ at a temperature T_z. Here, the dot denotes time derivatives and we introduced the Thomson optical depth, $\tau = \int \sigma_{\rm T} N_{\rm e} c\, {\rm d}t$, which is evaluated at the background level assuming the standard recombination history from `CosmoRec` [25]. The temperature variables are presented as $\theta_i = k_{\rm B}T_i/m_{\rm e}c^2$, with $\theta_z \propto T_z \propto (1+z)$. The Kompaneets operator is denoted by $\hat{\mathcal{K}}_x = x^{-2}\partial_x x^4 \partial_x + x^{-2}\partial_x x^4 A(x)$, with $A(x) = 1 + 2n_{\rm bb}(x)$, and $\Lambda = \Lambda(x, \theta_z)$ determines the photon production rate by double Compton (DC) and Bremsstrahlung (BR). These can be computed accurately using `BRpack` [26] and `DCpack` [27]. The electron temperature perturbation, $\Theta_{\rm eq}^{(0)} = \Delta T_{\rm e}^{(0)}/T_z$, and effective heating rate, $\dot{Q}^{(0)}$, are given by

$$\Theta_{\rm eq}^{(i)} = \frac{\int x^3 \Delta n_0^{(i)}\, w_y\, {\rm d}x}{4E_{n_{\rm bb}}} \qquad \text{and} \qquad \dot{Q}^{(0)} \equiv \frac{\dot{Q}_{\rm c}^{(0)}}{\rho_z}, \tag{2.2}$$

with y-weight factor $w_y = Y/G = xA(x) - 4 = x\frac{{\rm e}^x+1}{{\rm e}^x-1} - 4$ and where $E_f = \int x^3 f(x)\, {\rm d}x$ is the energy density integral of $f(x)$. The expression for $\Theta_{\rm eq}^{(i)}$ can be obtained by balancing Compton heating and cooling (Appendix C.1). The distortion sources from heating relate directly to $\dot{Q}_{\rm c}^{(0)}$ from collisions (see Appendix C.2). In equation (2.2), $\rho_z = \frac{8\pi h}{c^3}\left(\frac{k_{\rm B}T_z}{h}\right)^4 E_{n_{\rm bb}} \approx 0.261\,{\rm eV/cm^3}\,[T_0(1+z)/2.726\,{\rm K}]^4$ is the energy density of a blackbody at temperature T_z.

Usually, Eq. (2.1) is solved on a frequency-grid for the spectral distortion, $\Delta n^{(0)}(t,x)$, e.g., using `CosmoTherm` [28]. This task can become very time-consuming and certainly is not easily extendable to anisotropic distortions. The main result of paper I was to demonstrate that the problem can be approximately written as a matrix equation. The derivation uses the Ansatz $\Delta n^{(0)}(t,x) \approx \boldsymbol{B}(x)\cdot \boldsymbol{y}^{(0)}(t)$ to discretise the average photon spectrum. Here, $\boldsymbol{B} = (G(x), Y(x), Y_1(x), \dots, Y_N(n), M(x))^T$ denotes the computation basis, based on the standard distortion shapes, G, Y and M as well as the boosted signals, $Y_k = (1/4)^k \hat{O}_x^k Y$ with boost generator, $\hat{O}_x = -x\partial_x$. This then yields the evolution equation

$$\begin{aligned} \frac{\partial \boldsymbol{y}^{(0)}}{\partial t} &\approx \dot{\tau}\theta_z \left[M_{\rm K}\, \boldsymbol{y}^{(0)} + \boldsymbol{D}^{(0)}\right] + \frac{\dot{\boldsymbol{Q}}^{(0)}}{4}, \\ \boldsymbol{D}^{(0)} = \left(\gamma_T x_{\rm c}\, \mu^{(0)}, 0, 0, \dots, 0, -\gamma_N x_{\rm c}\, \mu^{(0)}\right)^T, &\qquad \dot{\boldsymbol{Q}}^{(0)} = \left(0, \dot{Q}^{(0)}, 0, \dots, 0, 0\right)^T. \end{aligned} \tag{2.3}$$

for the spectral parameters, $\boldsymbol{y}^{(0)}(t)$. Here, $x_{\rm c}$ is the critical frequency for photon emission, and the emission coefficients are $\gamma_T \approx 0.1387$ and $\gamma_N \approx 0.7769$. The Kompaneets mixing matrix, M_K describes the (photon number- and energy-conserving) rotation of the spectral parameters in each scattering, and is pre-computed for different basis sizes. From Eq. (2.3) one can already anticipate the main thermalisation terms; however, a few augmentations will become important as we show now. We also note, that for Eq. (2.3), the problem was linearised respect to the distortion. Possible sources of distortions from photon injection processes [e.g., 29, 30] were also neglected and cannot be readily treated within the proposed framework.

2.2 General statement of the problem in first order of perturbation theory

At first order in perturbation theory, the evolution equation for the anisotropic photon occupation number, $n^{(1)} = n^{(1)}(t, x, \boldsymbol{r}, \hat{\boldsymbol{\gamma}})$, at location $\boldsymbol{r}$ and in the direction $\hat{\boldsymbol{\gamma}}$ reads [e.g., 31–33]

$$\frac{\partial n^{(1)}}{\partial t} + \frac{c\hat{\boldsymbol{\gamma}}}{a} \cdot \nabla n^{(1)} + \hat{O}_x n^{(0)} \left(\frac{\partial \Phi^{(1)}}{\partial t} + \frac{c\hat{\boldsymbol{\gamma}}}{a} \cdot \nabla \Psi^{(1)} \right) = C^{(1)}[n]. \tag{2.4}$$

Here, we work with the metric perturbations in conformal Newtonian gauge in which the line element is $\mathrm{d}s^2 = a^2(-\mathrm{e}^{2\Psi}\,\mathrm{d}\eta^2 + \mathrm{e}^{2\Phi}\delta_{ij}\,\mathrm{d}x^i\,\mathrm{d}x^j)$, when neglecting vector and tensor perturbations; $C^{(1)}[n]$ denotes the rather complicated collision term [e.g., 24, 34–36], that accounts for the effect of Thomson scattering and thermalisation processes. Usually, one would neglect all distortions or thermalisation effects, such that $\hat{O}_x n^{(0)} \approx \hat{O}_x n_{\rm bb} \approx G(x)$. With the Ansatz $n^{(1)} \approx \Theta^{(1)}\,G(x)$, where $\Theta^{(1)} = \Delta T/T_0$ describes the fractional CMB temperature perturbation, this leads to the standard brightness temperature equations in the Thomson limit [31, 33, 37]:

$$\frac{\partial \Theta^{(1)}}{\partial t} + \frac{c\hat{\boldsymbol{\gamma}}}{a} \cdot \nabla \Theta^{(1)} + \frac{\partial \Phi^{(1)}}{\partial t} + \frac{c\hat{\boldsymbol{\gamma}}}{a} \cdot \nabla \Psi^{(1)} = \dot{\tau} \left[\Theta_0^{(1)} + \frac{1}{10}\,\Theta_2^{(1)} - \Theta^{(1)} + \beta^{(1)}\chi \right], \tag{2.5}$$

where $\Theta_\ell^{(1)}$ is the Legendre transform of the photon temperature field, $\Phi^{(1)}$ and $\Psi^{(1)}$ are the potential perturbations, $\beta^{(1)}$ the baryon speed, and $\chi = \hat{\boldsymbol{\beta}} \cdot \hat{\boldsymbol{\gamma}}$ is the direction cosine between the baryon velocity $\boldsymbol{\beta}$ and the photon direction, $\hat{\boldsymbol{\gamma}}$. Polarisation terms were not included here, but do not affect the main arguments presented below. For our numerical solutions, they are added back as usual [31]. Equation (2.5) will now be generalized to include the effect of a non-vanishing average distortion, spectral evolution, anisotropic heating and perturbed thermalisation. These effects lead to small corrections to the brightness temperature equations but give the leading order terms in the distortion hierarchy.

Accounting for all the thermalisation effects self-consistently is beyond the scope of this paper. However, if we assume that locally thermalisation is *only* mediated through the monopole part of the spectrum, the situation becomes more tractable. This is in fact well-motivated since for the anisotropies the much faster Thomson scattering process dominates, while Thomson terms are absent in the monopole, making thermalisation terms dominant there [24, 36]. In particular the monopole here is defined in the local inertial frame. We note, however, that changing frame would introduce similar transformations in the Thomson and Compton scattering terms, and the arguments of timescales would still hold true for the relevant pairings of non-monopole scattering terms.. We will furthermore not attempt to solve for higher order corrections to the CMB temperature anisotropy field, as this would entail a full treatment in second order perturbation theory [e.g., 35, 38, 39] including additional modifications to more correctly treat distortion terms.

To better appreciate the various new effects we will proceed in a step-by-step manner first only considering Thomson scattering terms, $\propto \dot{\tau}$. These will reveal the optimal basis given an average distortion and also already show the main effects in terms of distortion anisotropy generation. The spectral evolution across the various distortion types is then included using the modified Green's function treatment with the spectral mixing occurring only in the local monopole. The scattering efficiency is $\propto \dot{\tau}\theta_z$ and hence suppressed relative to the Thomson terms. However, at $z \gtrsim 10^4$, we expected spectral evolution to change the distortion signals in an observable way.

We give many of the derivation details in Appendix A. In addition, we will consider various limiting solution in Appendix B. In Appendix C, we furthermore explicitly carry out some of the approximations that are required to simplify the problem to the expressions given in the main text. We refer the interested reader to these sections for an in depth understanding. However, we hope that the result presented in this section can be mostly understood without the detailed (and sometimes cumbersome) derivations.

We also mention that the picture and interpretation of the results we develop below could potentially be cast into another form using alternative variables and reference frames. For example, one could use the local baryon frame to overcome some of the kinematic complications in the scattering process [36]. Similarly, log-temperature moments [40] might provide the means to separate some of the redshifting and scattering effects in a more transparent way. However, the methods applied here closely follow the standard approach for CMB anisotropy computations [31, 41], extending them in a minimal way to account for the evolution of the average spectrum. This is deemed to be a valuable step in understanding the rich physics of CMB spectral distortion anisotropies.

2.3 Effect of Doppler and potential terms in the Thomson limit

As the first and simplest scenario, let us consider the case where heating happened well before the recombination era and the average distortion is long frozen into its final state. We shall also assume that any thermalisation effects and energy exchange terms can be neglected. Thomson scattering isotropises the medium and leads to damping of the perturbations at small scales. This picture is valid for modes that become dynamic at $z \lesssim 10^4$. The photon evolution equation then reads

$$\frac{\partial n^{(1)}}{\partial t} + \frac{c\hat{\boldsymbol{\gamma}}}{a}\cdot\nabla n^{(1)} + \hat{O}_x n^{(0)}\left(\frac{\partial\Phi^{(1)}}{\partial t} + \frac{c\hat{\boldsymbol{\gamma}}}{a}\cdot\nabla\Psi^{(1)}\right) \approx \dot{\tau}\left[n_0^{(1)} + \frac{1}{10}n_2^{(1)} - n^{(1)} + \beta^{(1)}\chi\,\hat{O}_x n^{(0)}\right] \tag{2.6}$$

using the Thomson collision term, Eq. (A.1). Here, we defined $n_\ell(t, x, \boldsymbol{r}, \hat{\boldsymbol{\gamma}}) = \sum_m n_{\ell m}(t, x, \boldsymbol{r}) Y_{\ell m}(\hat{\boldsymbol{\gamma}})$ using the spherical harmonic coefficients of the photon occupation number, $n_{\ell m}(t, x, \boldsymbol{r})$. This definition implies $n^{(0)} \equiv n_0^{(0)} = n_{00}^{(0)} Y_{00} = n_{00}^{(0)}/\sqrt{4\pi}$ for the average spectrum, $n_0^{(0)} = \int n^{(0)} \frac{\mathrm{d}\Omega}{4\pi}$, where the solid angle is $\mathrm{d}\Omega = \mathrm{d}^2\hat{\boldsymbol{\gamma}} = \mathrm{d}\phi\,\mathrm{d}\chi$.

We now assume that the average spectrum is spectrally frozen, $n^{(0)}(x) = n_{\rm bb}(x) + \Delta n^{(0)}(x)$, where $\Delta n^{(0)}(x)$ is the departure from the blackbody, $n_{\rm bb}(x) = 1/[\mathrm{e}^x - 1]$ at a temperature T_z. We can then make the simple Ansatz $n^{(1)} \approx \Theta^{(1)} G(x) + \Sigma^{(1)} \hat{O}_x \Delta n^{(0)}$.[1] Assuming that $\hat{O}_x \Delta n^{(0)} \neq G(x)$, this results in two identical photon Boltzmann hierarchies for $\Theta^{(1)}$ and $\Sigma^{(1)}$ that are exactly like the standard equation for the temperature perturbations. These can be obtained by simply comparing coefficients of the two independent spectral functions or alternatively by taking the number and energy density moments of the Boltzmann equation (see Appendix B.1). However, in contrast to the temperature perturbations, at a given wavenumber k we start with the initial condition $\Sigma^{(1)} = 0$ such that distortion anisotropies are only sourced later when perturbations in the potentials and velocity field appear [24]. The transfer functions and related power spectra can be computed directly using any Einstein-Boltzmann code. Cross-correlations of the distortion and temperature fields thus probe Doppler and integrated Sachs-Wolfe (ISW) terms. This can result in a novel noise floor for tests of primordial non-Gaussianity from $\mu/y \times T$ correlations, as we highlight in paper III.

2.3.1 Preparing for more general spectral evolution

As explained above, in the Thomson limit the problem described by Eq. (2.6) can in principle be solved with only two independent variables, $\Theta^{(1)}$ and $\Sigma^{(1)}$ [see Eq. (B.4) for the Boltzmann equations], and the new spectral shape $\hat{O}_x \Delta n^{(0)}$ caused by boosts[2] of the average distortion. For example, if

---

[1] For the considered case, one could also directly use $n^{(1)} \approx \Sigma^{(1)} \hat{O}_x n^{(0)}$, but then the initial condition can not be as clearly separated and an extension to the time-dependent case is not as straightforward.

[2] In this work we will refer to $\hat{O}_x$ as the boost generator, because it appears in Lorentz transformations of a photon spectrum. The first order Doppler term $\dot{\tau}\beta^{(1)}\chi\,\hat{O}_x n^{(0)}$ in Eq. (2.6) has its origin in this. However, for the terms relating to the potentials, $\hat{O}_x$ physically appears due to gravitational redshifting, which we do not distinguish explicitly.

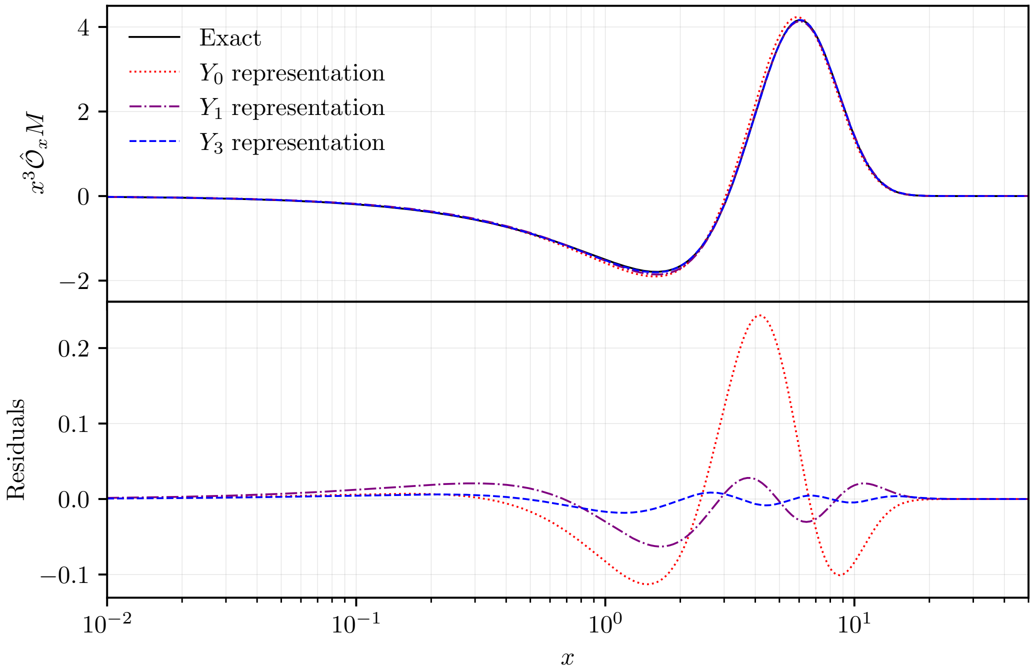


Figure 1: Representation of $\hat{O}_x M(x)$ using varying elements in the basis. Even the lowest order representation in terms of μ and y is highly accurate.

the average spectral distortion was a pure μ-distortion, the anisotropies would have the spectrum of $\hat{O}_x M(x)$. We could then observationally search for $G(x)$ and $\hat{O}_x M(x)$ and thereby extract information on the perturbations, $\Theta^{(1)} \propto \zeta$ and $\Sigma^{(1)} \propto \epsilon_\rho \zeta$.

In more general situations, when spectral evolution is also included, using the computational spectral basis we can write $\Delta n^{(0)} \approx \boldsymbol{y}^{(0)} \cdot \boldsymbol{B}$ and hence

$$\hat{O}_x \Delta n^{(0)} = \Theta^{(0)}[3G(x) + Y(x)] + 4 \sum_{k=1}^{N} y_{k-1}^{(0)}\, Y_k(x) + y_N^{(0)}\, \hat{O}_x Y_N(x) + \mu^{(0)}\, \hat{O}_x M(x), \tag{2.7}$$

where we used $\hat{O}_x G = 3G + Y$. The first two groups of terms directly fall back onto the original spectral basis, $\boldsymbol{B}$; however, $\hat{O}_x Y_N(x) = 4Y_{N+1}(x)$ and $\hat{O}_x M(x) = [M(x)/G(x)]\hat{O}_x G(x) - G(x)/x$ lie outside. As a simple fix, we could add these new spectral shapes to the basis and thereby keep the precision of the average distortion. However, once we consider Compton scattering effects, this approach becomes problematic, requiring a new truncation of the spectral hierarchy.

Instead of adding new spectra to the basis, we approximately represented $\hat{O}_x Y_N(x)$ and $\hat{O}_x M(x)$ within the old spectral basis.[3] The projection procedure is explained in detail in paper I. Just like for the Kompaneets operator, one has to ensure energy conservation in the calculation; photon number conservation is automatically built in, since Y_N and M do not carry number and hence $\hat{O}_x Y_N$ and $\hat{O}_x M$ do not either. It turns out that $\hat{O}_x M(x)$ is extremely well represented as a simple sum of μ and y. Carrying out the projections, we find $\hat{O}_x M(x) \approx 0.3736\, Y(x) + 1.9069\, M(x)$; however, even better precision can be achieved when including additional terms in $Y_k(x)$ (see Fig. 1). As for the Kompaneets operator (see paper I), the largest error is always made when representing $Y_{N+1}(x)$ with only terms up to $Y_N(x)$, as part of function space is omitted. However, less and less energy is carried in these contributions such that the precision does not suffer much once Y_5 or higher are included.

---

[3] We tried several alternatives but found none of them to be more beneficial. A more general study is left for the future.

With these comments, schematically we can therefore always write the representations

$$\hat{O}_x Y_N(x) = 4Y_{N+1}(x) \approx \boldsymbol{O}_N \cdot \boldsymbol{B}(x), \qquad \hat{O}_x M(x) \approx \boldsymbol{O}_\mu \cdot \boldsymbol{B}(x) \tag{2.8}$$

where $\boldsymbol{O}_N$ and $\boldsymbol{O}_\mu$ are the corresponding solution vectors in terms of $y, y_1, \dots, y_N, \mu$ (the projections of $G(x)$ never matter as $\hat{O}_x Y_k$ and $\hat{O}_x M$ do not carry photon number). This means we can express the boosted average spectrum as $\hat{O}_x n^{(0)} \approx \boldsymbol{b}^{(0)} \cdot \boldsymbol{B}$ with

$$\boldsymbol{b}^{(0)} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} + M_{\rm B}\,\boldsymbol{y}^{(0)} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 3 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & \cdots & 0 & 0 & O_{N,1} & O_{\mu,1} \\ 0 & 4 & 0 & 0 & \cdots & 0 & 0 & O_{N,2} & O_{\mu,2} \\ 0 & 0 & 4 & 0 & \cdots & 0 & 0 & O_{N,3} & O_{\mu,3} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & O_{N,N-2} & O_{\mu,N-2} \\ 0 & 0 & 0 & 0 & \cdots & 4 & 0 & O_{N,N-1} & O_{\mu,N-1} \\ 0 & 0 & 0 & 0 & \cdots & 0 & 4 & O_{N,N} & O_{\mu,N} \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & O_{N,N+1} & O_{\mu,N+1} \end{pmatrix} \begin{pmatrix} \Theta^{(0)} \\ y_0^{(0)} \\ y_1^{(0)} \\ y_2^{(0)} \\ \vdots \\ y_{N-2}^{(0)} \\ y_{N-1}^{(0)} \\ y_N^{(0)} \\ \mu^{(0)} \end{pmatrix}, \tag{2.9}$$

where we implicitly defined the boosting matrix, $M_{\rm B}$, noting that $\boldsymbol{B}$ has $N+3$ entries starting with $G(x)$ for $k=0$ and ending with $M(x)$ for $k=N+2$. Inserting $n^{(1)} = \boldsymbol{y}^{(1)} \cdot \boldsymbol{B}$ into Eq. (2.6) and rearranging terms we can then write the photon Boltzmann equation in the Thomson limit as

$$\frac{\partial \boldsymbol{y}^{(1)}}{\partial \eta} + \hat{\boldsymbol{\gamma}} \cdot \nabla \boldsymbol{y}^{(1)} \approx -\boldsymbol{b}_0^{(0)} \left(\frac{\partial \Phi^{(1)}}{\partial \eta} + \hat{\boldsymbol{\gamma}} \cdot \nabla \Psi^{(1)} \right) + \tau' \left[\boldsymbol{y}_0^{(1)} + \frac{1}{10} \boldsymbol{y}_2^{(1)} - \boldsymbol{y}^{(1)} + \beta^{(1)} \chi\, \boldsymbol{b}_0^{(0)} \right], \tag{2.10}$$

where we now converted to conformal time, $\eta = \int_0^t c\, {\rm d}t'/a$ with $\tau' = {\rm d}\tau / {\rm d}\eta = N_{\rm e} \sigma_{\rm T} a$, and also used $\boldsymbol{y}_\ell(t, \boldsymbol{r}, \hat{\boldsymbol{\gamma}}) = \sum_m \boldsymbol{y}_{\ell m}(t, \boldsymbol{r}) Y_{\ell m}(\hat{\boldsymbol{\gamma}})$ as before for n_ℓ. We furthermore made explicit that $\boldsymbol{b}^{(0)}$ only has a monopolar dependence, $\boldsymbol{b}^{(0)} = \boldsymbol{b}_0^{(0)} \equiv \int \boldsymbol{b}^{(0)} \frac{{\rm d}\Omega}{4\pi}$. For the example above, $\boldsymbol{b}_0^{(0)}$ was *time-independent*. In more general situations it becomes a function of time, where $\Delta \boldsymbol{b}_0^{(0)} = M_{\rm B}\, \boldsymbol{y}_0^{(0)}$ provides sources for distortion and temperature anisotropies. However, before we can treat this problem we need to include the effects of spectral evolution into the equations, as we discuss in Sect. 2.4.

2.3.2 Changing the background temperature

One of the simplest problems that one can already consider with Eq. (2.10) is a time-dependent change of the background temperature that leads to a departure from the standard $\propto (1+z)$ scaling. This limit describes average injection of G, which physically is hard to establish without efficient thermalisation terms [42], but provides some intuition. Speculative scenarios could relate to the decay of dark energy into photons [43, 44] or a treatment of energy release signals in the *instantaneous* *thermalisation* limit. It is also relevant to early treatments of temperature fluctuations at second order in perturbations [38], warranting the discussion presented here.

Using the zeroth order equation, Eq. (2.1), with $\Delta n^{(0)} = \Theta_0^{(0)} G$, and assuming that the change in the temperature is created by a time-dependent source term, $G\,S_T(t)$, we can easily verify that, as expected, this leaves the average spectrum unchanged:

$$\begin{aligned} G\,\frac{\partial\Theta_0^{(0)}}{\partial t} &= G\,S_T + \dot\tau\theta_z\left[\Theta_{\rm eq}^{(0)} Y + \hat{\mathcal{K}}_x \Delta n_0^{(0)}\right] - \dot\tau\,\frac{\Lambda(x,\theta_z)\,(1-{\rm e}^{-x})}{x^3}\Delta n_0^{(0)} + \dot\tau\,\frac{\Lambda(x,\theta_z)}{x^2}\,n_{\rm bb}\,\Theta_{\rm eq}^{(0)} \\ &= G\,S_T + \dot\tau\theta_z\,\Theta_0^{(0)}\left[Y + \hat{\mathcal{K}}_x G\right] - \dot\tau\,\frac{\Lambda(x,\theta_z)}{x^2}\,n_{\rm bb}\,\Theta_0^{(0)}\left[\frac{\Delta n_0^{(0)}}{\Theta_0^{(0)} G} - 1\right] \equiv G\,S_T. \end{aligned} \tag{2.11}$$

Here, we used $\Theta_{\rm eq}^{(0)} = \Theta_0^{(0)}$ and $(1-{\rm e}^{-x})/x = {\rm e}^{-x}({\rm e}^{x}-1)/x = n_{\rm bb}/G$ and assumed that the external source has a spectrum G. We also used $\hat{\mathcal{K}}_x G = -Y$, which causes all scattering terms to vanish, and therefore no new spectral shapes other than G appear in the average evolution.

We can now ask how the evolution of the average CMB spectrum propagates into the anisotropies. Knowing the solution of $\bar\Theta(t) = \Theta^{(0)}(t)$, from equation (2.10), we then find the two hierarchies

$$\frac{\partial\Theta^{(1)}}{\partial t} + \frac{c\hat\gamma}{a}\cdot\nabla\Theta^{(1)} + \bar b\left(\frac{\partial\Phi^{(1)}}{\partial t} + \frac{c\hat\gamma}{a}\cdot\nabla\Psi^{(1)}\right) \approx \dot\tau\left[\Theta_0^{(1)} + \frac{1}{10}\,\Theta_2^{(1)} - \Theta^{(1)} + \bar b\beta^{(1)}\chi\right]. \tag{2.12a}$$

$$\frac{\partial y^{(1)}}{\partial t} + \frac{c\hat\gamma}{a}\cdot\nabla y^{(1)} + \bar\Theta\left(\frac{\partial\Phi^{(1)}}{\partial t} + \frac{c\hat\gamma}{a}\cdot\nabla\Psi^{(1)}\right) \approx \dot\tau\left[y_0^{(1)} + \frac{1}{10}\,y_2^{(1)} - y^{(1)} + \bar\Theta\beta^{(1)}\chi\right]. \tag{2.12b}$$

with $\bar b(t) = (1 + 3\bar\Theta)$. This indicates that in addition to small modifications to the temperature anisotropies, apparent y-distortion anisotropies are sourced. The latter are due to the mismatch of the thermal spectrum of the initial anisotropies with respect to the new average blackbody at a temperature $\bar T = T_z(1+\bar\Theta)$, which generally does not scale like $\propto (1+z)$. In the absence of thermalisation effects, this indeed causes distortion anisotropies, as we illustrate now. However, since even a pure blackbody temperature fluctuation has a y-type spectral contribution at second order in Θ [40, 45, 46], to demonstrate that a distortion anisotropy is created we have to show that the photon number and energy densities in the anisotropies do not simply obey the blackbody relations, with one effective temperature describing the full spectrum.

To make progress, we therefore first ask the question how a spatially-varying blackbody spectrum changes when the average temperature is varied in a way that departs from the standard $\propto (1+z)$ scaling. For this, we write the expression

$$n = \frac{1}{\exp\left(h\nu/[k\bar T(1+\Theta)]\right) - 1} \equiv n_{\rm bb}\left(x/[(1+\bar\Theta)(1+\Theta)]\right), \tag{2.13}$$

where Θ describes the temperature anisotropies while $\bar\Theta$ is only time-dependent. At zeroth (no spatial terms) and first order in perturbations (only up to terms $\epsilon_\rho\zeta$) this implies

$$n^{(0)} = n_{\rm bb}(x) + \bar\Theta\,G(x) \tag{2.14a}$$

$$n^{(1)} = \Theta^{(1)}\,[G(x) + \bar\Theta\,\hat O_x G] = \Theta^{(1)}(1+3\bar\Theta)\,G(x) + \bar\Theta\Theta^{(1)}\,Y(x), \tag{2.14b}$$

where we neglected terms $O(\bar\Theta)^2$ and $O(\Theta^{(1)})^2$. We comment immediately on the consistency of this limit: terms $O(\bar\Theta)^2$ do not add any spatial effects and thus are merely higher order corrections to the average spectrum. Terms relating to $O(\Theta^{(1)})^2 \simeq O(\zeta)^2$ are second order in the primordial curvature perturbations, which we also do not consider here.

The expressions in Eq. (2.14) seems to suggest that a y-type distortion is sourced at first order, even if we started with a blackbody. However, the terms shown above are merely needed to precisely transition from a blackbody at the initial temperature to a new blackbody when $\bar{\Theta}$ varies [e.g., 45, 46]. This begs the question if this Ansatz solves the evolution equations in Eq. (2.12), which would show that no real distortion anisotropy is actually created?[4] Using $\mathcal{D}_t[X]$ from Eq. (B.2a), we have

$$\begin{aligned}\mathcal{D}_t[n^{(1)}] &= \mathcal{D}_t[\Theta^{(1)}]\,G(x) + \mathcal{D}_t[\bar{\Theta}\Theta^{(1)}]\hat{O}_x G = \mathcal{D}_t[\Theta^{(1)}]\,[G(x) + \bar{\Theta}\hat{O}_x G] + \mathcal{D}_t[\bar{\Theta}]\Theta^{(1)}\hat{O}_x G\\ &= \mathcal{D}_t[\Theta^{(1)}]\,\hat{O}_x n^{(0)} + \Theta^{(1)}\,S_T\,\hat{O}_x G,\end{aligned} \tag{2.15}$$

where $\hat{O}_x n^{(0)} \equiv G(x) + \bar{\Theta}\,\hat{O}_x G$ and $\mathcal{D}_t[\bar{\Theta}] = \dot{\bar{\Theta}} = S_T$. Put together with the rest of the Liouville operator [see Eq. (B.2) for relevant definitions], we then have

$$\mathcal{L}[n^{(1)}] = \mathcal{L}[\Theta^{(1)}]\,\hat{O}_x n^{(0)} + \Theta^{(1)}\,S_T\,\hat{O}_x G \equiv \hat{O}_x n^{(0)} C_T[\Theta^{(1)}] + C^{(1)}_{\rm therm}[n], \tag{2.16}$$

where we equated with the collision term, formally including both the Thomson terms, $C_T[\Theta^{(1)}]$, and thermalisation effects, $C^{(1)}_{\rm therm}[n]$. We also used $n^{(1)} \approx \Theta^{(1)}\,\hat{O}_x n^{(0)}$ to factor $\hat{O}_x n^{(0)}$ out of $C^{(1)}_T[n]$.

If we now only consider the Thomson terms and integrate photon number and energy density, with $\int x^2 \hat{O}_x n^{(0)}\,{\rm d}x \to 3N_z(1+3\bar{\Theta})$, $\int x^3 \hat{O}_x n^{(0)}\,{\rm d}x \to 4\rho_z(1+4\bar{\Theta})$, $\int x^2 \hat{O}_x G\,{\rm d}x \to 9N_z$ and $\int x^3 \hat{O}_x G\,{\rm d}x \to 16\rho_z$, we obtain the two equations

$$\begin{aligned}\int \mathcal{L}[n^{(1)}]\,x^2\,{\rm d}x \quad &\to \quad \mathcal{L}[\Theta^{(1)}] + 3\Theta^{(1)}\,S_T \approx C_T[\Theta^{(1)}],\\ \int \mathcal{L}[n^{(1)}]\,x^3\,{\rm d}x \quad &\to \quad \mathcal{L}[\Theta^{(1)}] + 4\Theta^{(1)}\,S_T \approx C_T[\Theta^{(1)}].\end{aligned} \tag{2.17}$$

Here, we used the number density, $N_z = \frac{8\pi}{c^3}\left(\frac{k_{\rm B}T_z}{h}\right)^3 N_{n_{\rm bb}} \approx 411.0\,{\rm cm}^{-3}\,[T_0(1+z)/2.725\,{\rm K}]^3$, of a blackbody at a temperature T_z and also neglected higher order terms in $\bar{\Theta}$. Since both equations have to be fulfilled, this shows that the spectrum cannot be a blackbody anymore without additional thermalisation processes. The reason is that $Y(x)$ does not carry photon number and only contributes to the second equation, leading to an extra source of $\Theta^{(1)}\,S_T$. The CMB anisotropies are thus distorted.

Since the only source spectra that are present in the Ansatz are G and Y, in the Thomson limit we need an extra independent y-parameter to describe the full spectrum. Adding an extra term $y_{\rm d} Y(x)$ to the Ansatz we previously used [i.e. using $n^{(1)} + y^{(1)}_{\rm d}\,Y(x)$ as our input], we then have

$$\begin{aligned}\mathcal{L}[n^{(1)}] &= \left(\mathcal{L}[\Theta^{(1)}]\,(1+3\bar{\Theta}) + 3\Theta^{(1)}\,S_T\right)G + \left(\bar{\Theta}\mathcal{L}[\Theta^{(1)}] + \Theta^{(1)}S_T + \mathcal{D}_t[y^{(1)}_{\rm d}]\right)Y\\ &\equiv C_T[\Theta^{(1)}](1+3\bar{\Theta})G + \left(\bar{\Theta}\,C_T[\Theta^{(1)}] + C_T[y^{(1)}_{\rm d}] - \beta^{(1)}\chi\right)Y + C^{(1)}_{\rm therm}[n].\end{aligned} \tag{2.18}$$

Here, we explicitly split the terms $\propto G$ and $\propto Y$, since we now wish to keep track of distortion contributions. By comparing coefficients we can find the modified system

$$\mathcal{L}[\Theta^{(1)}] + 3\Theta^{(1)}\,S_T \approx C_T[\Theta^{(1)}], \tag{2.19a}$$

$$\bar{\Theta}\mathcal{L}[\Theta^{(1)}] + \Theta^{(1)}\,S_T + \mathcal{D}_t[y^{(1)}_{\rm d}] \approx \bar{\Theta}\,C_T[\Theta^{(1)}] + C_T[y^{(1)}_{\rm d}] - \beta^{(1)}\chi, \tag{2.19b}$$

---

[4] In strictly speaking, we need to include a second order average source term with a y-type spectrum to leave the average blackbody unaffected (see Appendix B.2). This statement again emphasises that the appearance of terms $\propto Y(x)$ does not immediately imply a real distortion. However, we skipped this intermediate step in the derivation.

where we again neglected thermalisation effects. With $\bar{\Theta}\mathcal{L}[\Theta^{(1)}] \approx \bar{\Theta}\, C_T[\Theta^{(1)}]$, after dropping terms $\simeq O(\bar{\Theta})^2$, one then has the explicit evolution equations for $\Theta^{(1)}$ and $y_{\rm d}^{(1)}$

$$\frac{\partial \Theta^{(1)}}{\partial t} + \frac{c\hat{\gamma}}{a} \cdot \nabla \Theta^{(1)} + \frac{\partial \Phi^{(1)}}{\partial t} + \frac{c\hat{\gamma}}{a} \cdot \nabla \Psi^{(1)} \approx -3\Theta^{(1)} S_T + \dot{\tau}\left[\Theta_0^{(1)} + \frac{1}{10}\Theta_2^{(1)} - \Theta^{(1)} + \beta^{(1)}\chi\right], \quad (2.20a)$$

$$\frac{\partial y_{\rm d}^{(1)}}{\partial t} + \frac{c\hat{\gamma}}{a} \cdot \nabla y_{\rm d}^{(1)} \approx -\Theta^{(1)} S_T + \dot{\tau}\left[y_{\rm d,0}^{(1)} + \frac{1}{10} y_{\rm d,2}^{(1)} - y_{\rm d,0}^{(1)}\right]. \quad (2.20b)$$

These equations show that a non-vanishing y-parameter is created as perturbations mix through Thomson scattering and free streaming in the presence of average CMB temperature changes. As alluded to above, the reason is that the spectrum of the initial temperature perturbations disagrees with that of the evolving average blackbody. The change in the average blackbody temperature sources $\Theta^{(1)}$ and $y_{\rm d}^{(1)}$ perturbations but on average no photons are created in the fluctuating part, implying that the perturbed photon field is not consistent with that of pure blackbody temperature fluctuations: one cannot simply transform *all* y-type terms away by redefining the temperature perturbations.

This discussion also shows that the Ansatz in Eq. (2.14) can be recast in terms of the effective parameters $\Theta^{(1)}_{\rm eff} = \Theta^{(1)}(1 + 3\bar{\Theta})$ and $y^{(1)}_{\rm eff} = y^{(1)}_{\rm d} + \bar{\Theta}\Theta^{(1)}$. Indeed, using this redefinition with Eq. (2.20) we recover Eq. (2.12). This clearly separates the origin of the distortion anisotropies, identifying temperature raising y-contributions, $y^{(1)}_{\rm mix} = \bar{\Theta}\Theta^{(1)}$ from contributions that cannot thermalise, $y^{(1)}_{\rm d}$. These terms can in principle be distinguished from simple higher order corrections $\propto (\Theta^{(1)})^2$ or uncertainty $\propto \Delta T_0$ in the exact average present-day blackbody temperature $\propto \Delta T_0 \Theta^{(1)}$. For the former, this statement is evident since the related terms exhibit a different correlation structure. For the latter, the effect would be *time independent*, which does change the correlation structure. However, a more in depth discussion of these effects is beyond the scope of this work.

We mention that corrections $\propto \bar{\Theta}\,\zeta$ to the potentials and baryon velocity equations in principle also need to be added to the Boltzmann hierarchy. For the velocity terms, we will more explicitly discuss this in Sect. 2.6. However, here we are not interested in computing the *exact* corrections to the temperature power spectra caused by changes of the background temperature, nor will we consider cases when these terms become very important. For computing the physical y-parameter caused by changes in the average CMB temperature, we can simply use the old set of equations to solve for $\Theta^{(1)}$, since all corrections $\propto \bar{\Theta}\,\zeta$ recursively enter the evolution equation for $y^{(1)}$ at higher order in $\bar{\Theta}$. In the Thomson limit, we have thus completed the formulation of the problem and demonstrated that real y-type distortion anisotropies are created by changing the average temperature. These could potentially be used to test for departures from the standard CMB temperature-redshift relation; however, we leave a more detailed discussion to future work.

2.3.3 Instantaneous thermalisation

We can extend our discussion to the case where thermalisation is always instantaneous. At the background level we then have a changing temperature according to $\dot{\bar{\Theta}} = S_T$ relative to T_z like before. In the anisotropies, photon production and Compton scattering would always ensure that the spectrum of the fluctuating part thermalises under *energy conservation*. From the arguments leading up to Eq. (2.17), we can therefore directly write

$$\frac{\partial \Theta^{(1)}}{\partial t} + \frac{c\hat{\gamma}}{a} \cdot \nabla \Theta^{(1)} + \frac{\partial \Phi^{(1)}}{\partial t} + \frac{c\hat{\gamma}}{a} \cdot \nabla \Psi^{(1)} \approx -4\Theta^{(1)} S_T + \dot{\tau}\left[\Theta_0^{(1)} + \frac{1}{10}\Theta_2^{(1)} - \Theta^{(1)} + \beta^{(1)}\chi\right], \quad (2.21)$$

which with the parameter $\Theta^{(1)}_{\rm eff} = \Theta^{(1)}(1 + 4\bar{\Theta})$ can be cast into the equivalent form

$$\frac{\partial \Theta^{(1)}_{\rm eff}}{\partial t} + \frac{c\hat{\gamma}}{a} \cdot \nabla \Theta^{(1)}_{\rm eff} + \bar{b}\left(\frac{\partial \Phi^{(1)}}{\partial t} + \frac{c\hat{\gamma}}{a} \cdot \nabla \Psi^{(1)}\right) \approx \dot{\tau}\left[\Theta^{(1)}_{\rm eff,0} + \frac{1}{10}\,\Theta^{(1)}_{\rm eff,2} - \Theta^{(1)}_{\rm eff,0} + \bar{b}\beta^{(1)}\chi\right], \qquad (2.22)$$

with $\bar{b} = (1 + 4\bar{\Theta})$. The factor of 4 instead of 3 originates from the fact that the energy carried by the y-distortion in the previous case now appears in the blackbody part due to the assumption of instantaneous thermalisation.

Equation 2.22 indicates that one can simplify the computation by scaling all variables (also Φ, Ψ and β) to absorb the factor of $\bar{b} = (1 + 4\bar{\Theta})$. However, assuming a general time-dependence of $\bar{\Theta}$, new effects on the CMB power spectra would appear if this equation would be valid throughout the recombination era: the time-dependence of the modes would be modified by the evolution of the background temperature and a time-dependent rescaling of the *temperature contrast* would be folded into the shape of the CMB power spectra, an effect that can principally be captured with the above equations. Physically, this of course is an academic example to highlight how a coupling between modes and the background can be formulated. Indeed, this effect is an apparent 'super-horizon' effect, which becomes even more obvious when we think about times before BBN, where multiple phase transitions do increase the average blackbody temperature [e.g., 33, 47]. Due to quasi-instantaneous thermalisation, this leads to a rescaling of the temperature variables at *all scales*, leaving the adiabatic nature of the initial perturbations totally unchanged and merely modifying the initial conditions to account for this modification. Clearly, for modes entering the horizon during BBN or before, there is no tracer of the time-dependent effects that could be observed today, since all modes that witnessed this effect have dissipated away by Silk damping. At the normal CMB scales, we are left with the standard CMB anisotropy evolution (relative to the present-day higher temperature). However, in the primordial gravitational wave background, as an example, the traces of the phase transitions are in principle still visible [48, 49].

2.4 Effect of Compton scattering and photon production

Now that we understand how to account for the effect of boosting and Thomson scattering, we will next include Compton scattering and photon production. These only affect the spectral evolution of the local monopole and can be treated using our ODE representation of the Green's function. We will start by a simplistic treatment (Sect. 2.4.1) that just uses a perturbed version of the zeroth order thermalisation terms, Eq. (2.3). A more rigorous derivation (Sect. 2.4.2) shows that a few additional terms appear; however, the overall picture does not change crucially.

2.4.1 Simplistic inclusion of thermalisation terms

Starting from the description in Eq. (2.3), one can obtain a simple version for the thermalisation terms at first order in perturbation theory. Perturbing $\dot{\tau}$ and including the potential perturbation due to the local inertial frame transformation [e.g., following 35], one readily finds

$$\begin{aligned}
\left.\frac{\partial \boldsymbol{y}^{(1)}}{\partial t}\right|_{\rm therm} &\approx \dot{\tau}\theta_z\left[M_{\rm K}\,\boldsymbol{y}^{(1)} + \boldsymbol{D}^{(1)}\right] + \dot{\tau}\theta_z\left(\delta^{(1)}_{\rm b} + \Psi^{(1)}\right)\left[M_{\rm K}\,\boldsymbol{y}^{(0)} + \boldsymbol{D}^{(0)}\right] + \frac{3}{2}\,\dot{\tau}\theta_z\Theta^{(1)}_0\,\boldsymbol{D}^{(0)} + \frac{\dot{\boldsymbol{Q}}^{(1)}}{4},\\
\boldsymbol{D}^{(1)} &= \left(\gamma_T x_{\rm c}\,\mu^{(1)}, 0, 0, \dots, 0, -\gamma_N x_{\rm c}\,\mu^{(1)}\right)^T, \qquad \dot{\boldsymbol{Q}}^{(1)} = \left(0, \dot{Q}^{(1)}, 0, \dots, 0, 0\right)^T \qquad (2.23)\\
\dot{Q}^{(1)} &= \frac{\dot{Q}^{(1)}_{\rm c}}{\rho_z} + \Psi^{(1)}\,\frac{\dot{Q}^{(0)}_{\rm c}}{\rho_z}.
\end{aligned}$$

While the equation above has been obtained in a simplistic manner, it actually turns out that even a more rigorous derivation does not change the result that much. The first group of terms simply describes the spectral evolution of the first order distortion parameters. The second accounts for corrections with respect to the average evolution from perturbations in the electron density and potentials. Here, we used $\dot{\tau}^{(1)}/\dot{\tau} \approx \delta_{\rm b}^{(1)}$, assuming that we are far from the recombination era, such that perturbed recombination effects [e.g., 35, 50, 51] can be omitted. The perturbed heating rate similarly includes perturbations in the first order heating term from collisions in the local inertial frame, $\dot{Q}_{\rm c}^{(1)}$, but also the effect of potentials on the zeroth order term.

However, the term $\frac{3}{2}\,\dot{\tau}\theta_z\Theta_0^{(1)}\,\boldsymbol{D}^{(0)}$ deserves a bit more explanation. It stems from perturbing the critical frequency, $x_{\rm c}$, with respect to the local photon temperature. We assumed that only DC is relevant for the emission processes, such that the photon emissivity is $\propto \theta_\gamma x_{\rm c}(\theta_\gamma) \simeq \theta_\gamma^{3/2}$. Inserting $\theta_\gamma = \theta_z(1+\Theta_0)$ then yields $[\theta_\gamma x_{\rm c}(\theta_\gamma)]^{(1)} \approx \theta_z x_{\rm c}(\theta_z)\,(3/2)\,\Theta_0^{(1)}$. This term thus relates to perturbed emission effects and modulates the zeroth order term, $\boldsymbol{D}^{(0)}$. However, as we show below the coefficient of this term indeed changes to unity when a more careful account for modifications to the local Compton and DC rate is carried out.

2.4.2 More rigorous treatment

In this section, we now obtain the thermalisation terms in a more rigorous manner. For this we have to follow a few steps as outlined in Appendix A and C. After writing the full photon collision term in Appendix A, one has to obtain the electron temperature in the given distorted radiation field. For this we make use of the result of [35] and include the effect of Compton scattering (Appendix C.1). While thermalisation terms are relevant, the electron temperature will always be extremely close to the Compton equilibrium temperature, which greatly simplifies the problem and leads to an effective heating term in the photon field as we demonstrate in Appendix C.2. The main result of that section is Eq. (C.11) for the Compton terms and heating sources. This expression only neglects one stimulated scattering correction, as explained in that section, but otherwise is consistent at order $O(\epsilon_\rho\zeta)$.

Translating Eq. (C.11) into matrix form using our spectral basis, we find the generalized Kompaneets terms at first order in perturbations

$$\begin{aligned}
\left.\frac{\partial \boldsymbol{y}^{(1)}}{\partial t}\right|_{\rm K} &\approx \dot{\tau}\theta_z M_{\rm K}\,\boldsymbol{y}^{(1)} + \dot{\tau}\theta_z\left(\delta_{\rm b}^{(1)} + \Psi^{(1)}\right) M_{\rm K}\,\boldsymbol{y}^{(0)} + \frac{\dot{\boldsymbol{Q}}^{(1)}}{4} + \dot{\tau}\theta_z\,\Theta_0^{(1)}\left[M_{\rm D}\,\boldsymbol{y}^{(0)} - \boldsymbol{S}^{(0)}\right],\\
\boldsymbol{S}^{(0)} &= \left(0, \delta_{\gamma,0}^{(0)} + 4\Theta_{\rm e}^{(0)}, -4\Theta_{\rm e}^{(0)}, \dots, 0, 0\right)^T,
\end{aligned} \tag{2.24}$$

where $M_{\rm D} = (M_{\rm B} - 3I)M_{\rm B}$ is the *Doppler matrix* (see Appendix C.2.1) and $\Theta_{\rm e}^{(0)} = \Theta_{\rm eq}^{(0)} + \dot{Q}_{\rm c}^{(0)}/[4\dot{\tau}\theta_z\rho_z]$ is the average electron temperature. We also used $\delta_{\gamma,0}^{(0)} = E_{\Delta n_0^{(0)}}/E_{n_{\rm bb}} = 4\Theta_0^{(0)} + 4\sum_n y_{0,n}^{(0)} + \epsilon_M\mu_0^{(0)}$ for the photon energy density perturbation. Aside from the last group of terms, all the other are already present in Eq. (2.23). As explained in Appendix C.2, the latter do not add any energy to the system, but merely lead to a more minor change in the spectral distortion evolution.

In Appendix C.3 we carry out a careful derivation of the photon production terms. It turns out that at first order in perturbations, a correction to the DC emissivity causes a modification of the related temperature correction term, changing $\frac{3}{2}\,\dot{\tau}\theta_z\Theta_0^{(1)}\,\boldsymbol{D}^{(0)} \to \dot{\tau}\theta_z\Theta_0^{(1)}\,\boldsymbol{D}^{(0)}$ (see Appendix C.3.2). The cause of this change is the precise balance between Compton and DC scattering terms, as explained there. It is expected that the coefficient can change depending on which approximation for the DC emissivity is actually used. In paper III, we will see that this modification is not expected to be as severe, but additional work may be needed. Put together we then have the thermalisation terms

$$
\begin{aligned}
\left.\frac{\partial \boldsymbol{y}^{(1)}}{\partial t}\right|_{\rm therm} &\approx \dot{\tau}\theta_z \left[M_{\rm K}\, \boldsymbol{y}^{(1)} + \boldsymbol{D}^{(1)}\right] + \dot{\tau}\theta_z \left(\delta_{\rm b}^{(1)} + \Psi^{(1)}\right)\left[M_{\rm K}\, \boldsymbol{y}^{(0)} + \boldsymbol{D}^{(0)}\right] + \dot{\tau}\theta_z \Theta_0^{(1)}\, \boldsymbol{D}^{(0)} + \frac{\dot{\boldsymbol{Q}}^{(1)}}{4} \\
&\qquad + \dot{\tau}\theta_z\, \Theta_0^{(1)} \left[M_{\rm D}\, \boldsymbol{y}^{(0)} - \boldsymbol{S}^{(0)}\right], \\
\boldsymbol{D}^{(1)} &= \left(\gamma_T x_{\rm c}\, \mu^{(1)}, 0, 0, \dots, 0, -\gamma_N x_{\rm c}\, \mu^{(1)}\right)^T, \qquad \dot{\boldsymbol{Q}}^{(1)} = \left(0, \dot{Q}^{(1)}, 0, \dots, 0, 0\right)^T \\
\boldsymbol{S}^{(0)} &= \left(0, \delta_{\gamma,0}^{(0)} + 4\Theta_{\rm e}^{(0)}, -4\Theta_{\rm e}^{(0)}, \dots, 0, 0\right)^T, \\
\dot{Q}^{(1)} &= \frac{\dot{Q}_{\rm c}^{(1)}}{\rho_z} + \Psi^{(1)} \frac{\dot{Q}_{\rm c}^{(0)}}{\rho_z}.
\end{aligned}
\tag{2.25}
$$

As explained in Appendix A.2.1, we neglected kinematic corrections to the perturbed thermalisation terms, as these contribute at order $\simeq \beta\,\dot{\tau}\,\theta_z$ and are therefore smaller than the boosting terms $\propto \beta\,\dot{\tau}$ in the Thomson contributions. We do not expect this to modify the main conclusions although some details might differ during the μ-era. Alternatively, one could perform the computation in the baryon frame to avoid this complication of the calculation [see 36, for discussion], however, this is beyond the scope of this work.

2.5 Final evolution equation

We are now in the position to write the full evolution equation in terms of the spectral basis $\boldsymbol{y}^{(0)}$ and $\boldsymbol{y}^{(1)}$. Together with Eq. (2.9), (2.10) and Eq. (2.25) we find

$$
\frac{\partial \boldsymbol{y}_0^{(0)}}{\partial \eta} = \tau' \theta_z \left[M_{\rm K}\, \boldsymbol{y}_0^{(0)} + \boldsymbol{D}_0^{(0)}\right] + \frac{\boldsymbol{Q}'^{(0)}}{4}, \tag{2.26a}
$$

$$
\begin{aligned}
\frac{\partial \boldsymbol{y}^{(1)}}{\partial \eta} + \hat{\boldsymbol{\gamma}} \cdot \nabla \boldsymbol{y}^{(1)} &= -\boldsymbol{b}_0^{(0)} \left(\frac{\partial \Phi^{(1)}}{\partial \eta} + \hat{\boldsymbol{\gamma}} \cdot \nabla \Psi^{(1)}\right) + \tau' \left[\boldsymbol{y}_0^{(1)} + \frac{1}{10} \boldsymbol{y}_2^{(1)} - \boldsymbol{y}^{(1)} + \beta^{(1)} \chi\, \boldsymbol{b}_0^{(0)}\right] + \frac{\boldsymbol{Q}'^{(1)}}{4} \\
&+ \tau' \theta_z \left\{M_{\rm K}\, \boldsymbol{y}_0^{(1)} + \boldsymbol{D}_0^{(1)} + \left[\delta_{\rm b}^{(1)} + \Psi^{(1)}\right]\left(M_{\rm K}\, \boldsymbol{y}_0^{(0)} + \boldsymbol{D}_0^{(0)}\right) + \Theta_0^{(1)} \left(\boldsymbol{D}_0^{(0)} + M_{\rm D}\, \boldsymbol{y}^{(0)} - \boldsymbol{S}^{(0)}\right)\right\},
\end{aligned}
\tag{2.26b}
$$

where we made it explicit that the zeroth order solution is only a monopole and converted to conformal time, η. The first line accounts for all terms in the Thomson limit and the effect of anisotropic heating. The last row describes the spectral evolution, with $M_{\rm K}\, \boldsymbol{y}_0^{(1)} + \boldsymbol{D}_0^{(1)}$ determining the main terms and the other being related to perturbed thermalisation effects. Before we convert this into a photon distortion parameter hierarchy in Fourier space, let us briefly discuss the expected effect of distortions on the other Einstein-Boltzmann equations.

2.6 Effect of distortion on the other perturbation equations

While we have completed our reformulation of the photon evolution equation in the presence of spectral distortions, we still have to consider how the other perturbation equations might be modified. As an example, let us update the momentum exchange equation with the baryons. For this we have to compute the integral $\frac{1}{2} \int x^3 \chi C^{(1)}[n]\, {\rm d}x\, {\rm d}\chi$. We shall neglect any small corrections from photon emission and absorption terms. Since the Compton terms all only act on the monopole[5], these also do not contribute and one is only left with the Thomson scattering terms. The change of the baryon momentum by scattering with photons at first order in perturbations then is [33]

---

[5] This statement changes if kinematic corrections are taken into account, but the correction is suppressed by a factor of $\theta_z \ll 1$ in comparison to the Thomson terms.

$$
\begin{aligned}
\left.\frac{\partial \beta^{(1)}}{\partial \eta}\right|_{\mathrm{T}} &\approx -\tau' \,\frac{\rho_z}{\rho_{\mathrm{b}}}\,\frac{1}{2}\int \frac{x^3\chi}{E_{n_{\mathrm{bb}}}}\left[n_0^{(1)}+\frac{1}{10}\,n_2^{(1)}-n^{(1)}+\beta^{(1)}\chi\,\hat{O}_x n_0^{(0)}\right]\mathrm{d}x\,\mathrm{d}\chi \\
&= \frac{\tau'}{\mathbb{i}}\,\frac{\rho_z}{\rho_{\mathrm{b}}}\int \frac{x^3\mathrm{d}x}{E_{n_{\mathrm{bb}}}}\left[\tilde{n}_1^{(1)}-\mathbb{i}\,\frac{\beta^{(1)}}{3}\,\hat{O}_x\tilde{n}_0^{(0)}\right] \quad\longrightarrow \\
\left.\frac{\partial \tilde{\beta}^{(1)}}{\partial \eta}\right|_{\mathrm{T}} &= \tau'\,\frac{4\rho_z}{3\rho_{\mathrm{b}}}\left[3\left(\tilde{\Theta}_1^{(1)}+\sum_{k=0}^{N}\tilde{y}_{k,1}^{(1)}+\frac{\epsilon_M}{4}\,\tilde{\mu}_1^{(1)}\right)-\tilde{\beta}^{(1)}\left(1+4\tilde{\Theta}_0^{(0)}+4\sum_{k=0}^{N}\tilde{y}_{k,0}^{(0)}+\epsilon_M\,\tilde{\mu}_0^{(0)}\right)\right]
\end{aligned}
\tag{2.27}
$$

with $\tilde{\beta}^{(1)} \equiv \mathbb{i}\beta^{(1)}$ and $\tilde{f}_\ell(\eta, x, \boldsymbol{r}) = \frac{\mathbb{i}^\ell}{2}\int P_\ell(\chi) f(\eta, x, \chi, \boldsymbol{r})\,\mathrm{d}\chi = \mathbb{i}^\ell f_{\ell 0}(\eta, x, \boldsymbol{r})/\sqrt{4\pi\,(2\ell+1)}$ based on the Legendre polynomials, $P_\ell(\chi)$. Also, ρ_z and ρ_{b} are the background level quantities (without any distortion effects included).

The presence of distortions modifies the momentum exchange by a small amount. For consistency, these terms should be included in the hierarchies presented below. Here, $\rho_z(1+\delta_{\gamma,0}^{(0)})$ is the total energy density of the average CMB with $\delta_{\gamma,0}^{(0)} = 4\tilde{\Theta}_0^{(0)}+4\sum_{k=0}^{N}\tilde{y}_{k,0}^{(0)}+\epsilon_M\,\tilde{\mu}_0^{(0)}$. Similarly, the momentum carried by the photon dipole is modified to $\rho_z\,(\tilde{\Theta}_1^{(1)}+\delta_{\gamma,1}^{(1)})$ with $\delta_{\gamma,1}^{(1)} = \sum_{k=0}^{N}\tilde{y}_{k,1}^{(1)}+\frac{\epsilon_M}{4}\,\tilde{\mu}_1^{(1)}$. In the tight coupling limit, one then has

$$
\tilde{\beta}_{\mathrm{tc}}^{(1)} \approx \frac{3\left(\tilde{\Theta}_1^{(1)}+\delta_{\gamma,1}^{(1)}\right)}{1+\delta_{\gamma,0}^{(0)}} \approx \frac{3\left(\tilde{\Theta}_1^{(1)}+\sum_{k=0}^{N}\tilde{y}_{k,1}^{(1)}+\frac{\epsilon_M}{4}\,\tilde{\mu}_1^{(1)}\right)}{1+4\tilde{\Theta}_0^{(0)}+4\sum_{k=0}^{N}\tilde{y}_{k,0}^{(0)}+\epsilon_M\,\tilde{\mu}_0^{(0)}} \approx 3\tilde{\Theta}_1^{(1)}+O(\tilde{\Theta}_1^{(1)}\Delta n^{(0)}). \tag{2.28}
$$

However, in practice the terms $\delta_{\gamma,0}^{(0)}$ and $\delta_{\gamma,1}^{(0)}$ are small corrections to the evolution equations of the standard variables unless the average distortion amplitude becomes close to unity, which is ruled out by *COBE/FIRAS*. In a similar way, the equations for the potentials and neutrinos as well as temperature polarisation states see negligible effects from the presence of distortions unless a full second order treatment is attempted. For our computation below we shall leave all the first order equations unchanged and only add the new equations describing the distortion anisotropies. Unless anisotropic heating effects are explicitly considered, this means that distortion perturbations are directly driven by the standard perturbations known from the evolution of the CMB temperature with no direct distortion anisotropy sources from the distortions themselves, the latter being $O(\Delta n^{(0)})^2$.

2.7 Expression in Fourier space

To obtain the SD transfer functions we now carry out the final step of going to Fourier space. This allows us to obtain the photon transfer functions using a treatment that is similar to that of the standard perturbation equations [52, 53]. All the variables appearing above are functions of $X(\eta, \chi, \boldsymbol{r})$. Going to Fourier space and using the expansion, $X(\eta, \chi, \boldsymbol{r}) = \sum_\ell (2\ell+1)\,(-\mathbb{i})^\ell \tilde{X}_\ell(\eta, \boldsymbol{r})\,P_\ell(\chi)$ in terms of the Legendre polynomials, we then have [33, 37]

$$
\frac{\partial X}{\partial \eta}+\hat{\boldsymbol{\gamma}}\cdot\nabla X \qquad\longrightarrow\qquad \frac{\partial X}{\partial \eta}+\mathbb{i}k\chi X \qquad\longrightarrow\qquad \frac{\partial \tilde{X}_\ell}{\partial \eta}+k\left(\frac{\ell+1}{2\ell+1}\tilde{X}_{\ell+1}-\frac{\ell}{2\ell+1}\tilde{X}_{\ell-1}\right), \tag{2.29}
$$

where k denotes the wavenumber of the mode. The background evolution equation in Eq. (2.26) needs no further modification. Carrying out the Fourier and Legendre transformations of all remaining terms, from Eq. (2.26) we then find the final photon hierarchy

$$\begin{aligned}
\frac{\partial \boldsymbol{y}_0^{(0)}}{\partial \eta} &= \tau' \theta_z \left[M_{\rm K}\, \boldsymbol{y}_0^{(0)} + \boldsymbol{D}_0^{(0)} \right] + \frac{\boldsymbol{Q}'^{(0)}}{4}, && \text{(2.30a)} \\
\frac{\partial \tilde{\boldsymbol{y}}_0^{(1)}}{\partial \eta} &= -k\, \tilde{\boldsymbol{y}}_1^{(1)} - \frac{\partial \tilde{\Phi}^{(1)}}{\partial \eta} \boldsymbol{b}_0^{(0)} + \frac{\boldsymbol{Q}'^{(1)}}{4} && \text{(2.30b)} \\
&\qquad + \tau' \theta_z \left\{ M_{\rm K}\, \tilde{\boldsymbol{y}}_0^{(1)} + \boldsymbol{D}_0^{(1)} + \left[\tilde{\delta}_{\rm b}^{(1)} + \tilde{\Psi}^{(1)} \right] \left(M_{\rm K}\, \boldsymbol{y}_0^{(0)} + \boldsymbol{D}_0^{(0)} \right) + \tilde{\Theta}_0^{(1)} \left(\boldsymbol{D}_0^{(0)} + M_{\rm D}\, \boldsymbol{y}^{(0)} - \boldsymbol{S}^{(0)} \right) \right\}, \\
\frac{\partial \tilde{\boldsymbol{y}}_1^{(1)}}{\partial \eta} &= k \left(\frac{1}{3} \tilde{\boldsymbol{y}}_0 - \frac{2}{3} \tilde{\boldsymbol{y}}_2 \right) + \frac{k}{3} \tilde{\Psi}^{(1)}\, \boldsymbol{b}_0^{(0)} - \tau' \left[\tilde{\boldsymbol{y}}_1^{(1)} - \frac{\tilde{\beta}^{(1)}}{3} \boldsymbol{b}_0^{(0)} \right], && \text{(2.30c)} \\
\frac{\partial \tilde{\boldsymbol{y}}_2^{(1)}}{\partial \eta} &= k \left(\frac{2}{5} \tilde{\boldsymbol{y}}_1^{(1)} - \frac{3}{5} \tilde{\boldsymbol{y}}_3^{(1)} \right) - \frac{9}{10}\, \tau'\, \tilde{\boldsymbol{y}}_2^{(1)}, && \text{(2.30d)} \\
\frac{\partial \tilde{\boldsymbol{y}}_{\ell \geq 3}^{(1)}}{\partial \eta} &= k \left(\frac{\ell}{2\ell + 1} \tilde{\boldsymbol{y}}_{\ell - 1} - \frac{\ell + 1}{2\ell + 1} \tilde{\boldsymbol{y}}_{\ell + 1} \right) - \tau' \tilde{\boldsymbol{y}}_\ell^{(1)}. && \text{(2.30e)}
\end{aligned}$$

Aside from the new terms in the monopole equations and the generalization of the potential and Doppler sources in terms of the distortion parameters this system is identical to the standard brightness equations [31, 33]. Setting the terms $\tau' \theta_z$ and $\boldsymbol{Q}'$ to zero and using $\boldsymbol{b}_0^{(0)} = (1, 0, \dots, 0)$ these are identically recovered.[6]

We note, however, that here we neglected corrections from polarisation terms, which affect the quadrupole equation [37, 54, 55]. For the standard temperature terms we can include these as usual, but no attempt to correct polarisation effects for the distortion evolution are made here. This means that the damping scales of the temperature and distortion anisotropies will differ slightly. For the distortions, the Thomson scattering terms will be identical [56]; however, additional effects in the thermalisation terms will have to be studied more carefully, a task that is left to future work but should be possible, e.g., by starting from [36, 39].

2.8 Line of sight integration and power spectra

To close our discussion of the theoretical aspects, we briefly present the line of sight approach for obtaining the final signal power spectra. In principle the solutions to the photon hierarchy in Eq. (2.30) are enough to compute the power spectra and cross-power spectra of two observables X and Y:

$$C_\ell^{XY}(\eta) = \frac{2}{\pi} \int k^2 \, {\rm d}k \, P(k)\, \hat{X}_\ell(\eta, k)\, \hat{Y}_\ell(\eta, k) \qquad \text{(2.31)}$$

once the corresponding transfer functions, in this context indicated by the hat, are obtained. However, this brute force approach becomes numerically challenging, as known from the standard CMB anisotropies [57]. To simplify matters, we start by Fourier transforming Eq. (2.26b), which yields

$$\begin{aligned}
\frac{\partial \boldsymbol{y}^{(1)}}{\partial \eta} + \mathbb{i} k \chi\, \boldsymbol{y}^{(1)} + \tau' \boldsymbol{y}^{(1)} &= \boldsymbol{S}_{\rm LOS}, && \text{(2.32a)} \\
\boldsymbol{S}_{\rm LOS} \equiv - \left(\frac{\partial \Phi^{(1)}}{\partial \eta} + \mathbb{i} k \chi \Psi^{(1)} \right) \boldsymbol{b}_0^{(0)} &+ \tau' \left[\boldsymbol{y}_0^{(1)} + \frac{1}{10}\, \boldsymbol{y}_2^{(1)} + \beta^{(1)} \chi\, \boldsymbol{b}_0^{(0)} \right] + \frac{\boldsymbol{Q}'^{(1)}}{4} && \text{(2.32b)} \\
&+ \tau' \theta_z \left\{ M_{\rm K}\, \boldsymbol{y}_0^{(1)} + \boldsymbol{D}_0^{(1)} + \left[\delta_{\rm b}^{(1)} + \Psi^{(1)} \right] \left(M_{\rm K}\, \boldsymbol{y}_0^{(0)} + \boldsymbol{D}_0^{(0)} \right) + \Theta_0^{(1)} \left(\boldsymbol{D}_0^{(0)} + M_{\rm D}\, \boldsymbol{y}^{(0)} - \boldsymbol{S}^{(0)} \right) \right\},
\end{aligned}$$

---

[6] In comparison to [37] we use $\Theta_\ell^{\rm Hu} = (2\ell + 1)\Theta_\ell$, which also is the definition used in [31]. For Φ and Ψ, we follow the sign convention of [37], which means we have $\Phi^{\rm Ma} = -\Phi$ as defined in [31].

where now all variables are functions $f^{(1)} = f^{(1)}(\eta, \chi, k)$. Realizing that the only real differences with respect to the temperature only case are that we now are dealing with a solution vector and a vector for the sources, following the standard steps (see Appendix D) we can directly write down the solution for $\tilde{\boldsymbol{y}}_\ell^{(1)}(\eta_f, k)$ at the final conformal time, η_f, as

$$\tilde{\boldsymbol{y}}_\ell^{(1)}(\eta_f, k) = \int_0^{\eta_f} \mathrm{d}\eta \, g(\eta) \, \tilde{\boldsymbol{S}}_\ell(\eta, \eta_f, k), \tag{2.33}$$

$$\tilde{\boldsymbol{S}}_\ell(\eta, \eta_f, k) = \left[\tilde{\boldsymbol{y}}_0^{(1)} + \tilde{\Psi}^{(1)} \boldsymbol{b}_0^{(0)} + \left(\frac{\partial \tilde{\Psi}^{(1)}}{\partial \eta} - \frac{\partial \tilde{\Phi}^{(1)}}{\partial \eta}\right) \frac{\boldsymbol{b}_0^{(0)}}{\tau'}\right] j_\ell(k\Delta\eta) + \tilde{\beta}^{(1)} \boldsymbol{b}_0^{(0)} \, j_\ell^{(1,0)}(k\Delta\eta) + \frac{\tilde{\boldsymbol{y}}_2^{(1)}}{2} \, j_\ell^{(2,0)}(k\Delta\eta)$$

$$+ \left\{\theta_z \Big[M_{\rm K} \, \tilde{\boldsymbol{y}}_0^{(1)} + \tilde{\boldsymbol{D}}_0^{(1)} + \left[\tilde{\delta}_{\rm b}^{(1)} + \tilde{\Psi}^{(1)}\right] \left(M_{\rm K} \, \boldsymbol{y}_0^{(0)} + \boldsymbol{D}_0^{(0)}\right) + \tilde{\Theta}_0^{(1)} \left(\boldsymbol{D}_0^{(0)} + M_{\rm D} \, \boldsymbol{y}^{(0)} - \boldsymbol{S}^{(0)}\right)\Big] + \frac{\boldsymbol{Q}'^{(1)}}{4\tau'}\right\} j_\ell(k\Delta\eta).$$

Here, we introduced the Thomson visibility function, $g(\eta) = \tau' \, \mathrm{e}^{-\tau_{\rm b}} = \partial_\eta \mathrm{e}^{-\tau_{\rm b}}$ with $\tau_{\rm b} = \tau(\eta_f) - \tau(\eta)$ and the Thomson optical depth $\tau = \int_0^\eta \tau'(\eta') \, \mathrm{d}\eta'$. We also defined $\Delta\eta = \eta_f - \eta$ for convenience. The functions $j_\ell^{(a,b)}(x)$ are based on the usual spherical Bessel functions $j_\ell(x)$ [58]. Concretely we give $j_\ell^{(1,0)}(x) = j_\ell'(x) = \partial_x j_\ell(x)$ and $j_\ell^{(2,0)}(x) = \frac{1}{2}\left[3 j_\ell''(x) + j_\ell(x)\right]$ as in [37].

In comparison to the standard line of sight approach, we have the extra source terms coming firstly from the spectral mixing by Compton scattering and perturbed thermalisation effects ($\propto \theta_z$) and secondly from anisotropic heating ($\propto \boldsymbol{Q}'^{(1)}$). In addition, we have a more general Doppler and potential source vector ($\propto \boldsymbol{b}_0^{(0)}$), which in general is time-dependent.

2.8.1 Reduction to the experimental basis

As explained in Sect. 3 of paper I, with an experiment in mind we can reduce the number of spectral parameters by going from the computation basis to the residual distortion representation. From the results for the transfer functions, this can be achieved by applying a matrix L to the solution vector $\boldsymbol{y}^{(1)}$. We can then plot the transfer functions for a more limited number of parameters, e.g., Θ, y, r_1 and μ, without loosing much information. In a similar manner we can simplify the computation of power spectra to those of the parameters in the experimental basis. Here is it best to directly apply the rotation to the source vector in the line-of-sight solution, Eq. (2.33). This reduces the computational burden, since the power spectra need to be computed for a reduced number of variables.

We comment that directly applying the projection on the limited observation basis before even computing the transfer function is not an optimal choice as in this case energy conservation is not guaranteed to the same level of precision as with the computation basis. Even if this would further reduce the computational burden, we thus do not recommend such an approach.

2.9 Basic expectations for the evolution of distortion anisotropies

While it is difficult to anticipate the detailed behavior of the distortion transfer functions by looking at the system in Eq. (2.30), we can already understand some of the most important features. Firstly, without average distortions, non-standard evolution of the average temperature (see Sect. 2.3.2) or anisotropic heating *no* distortion anisotropies are generated at first order even in the perturbed universe [24]. This limit is evident without further explanation, as in this case the system simply becomes identical to the standard temperature anisotropies at first order in perturbation theory.

Secondly, photon emission and Comptonisation effects $\propto \theta_z \simeq 4.6 \times 10^{-10}(1+z)$ can only be important at $z \gtrsim 10^4$, like for the average distortion evolution. At later times, the spectral evolution is mostly *frozen* and only direct distortion anisotropy sources from boosting and potentials ($\propto \hat{O}_x \Delta n^{(0)}$) or anisotropic heating ($\propto \dot{Q}^{(1)} Y$) can create additional distortion anisotropies at significant levels. We note that if isotropic heating is present (e.g., by particle decay) it is naturally expected that anisotropic heating occurs unless the heating mechanism is detached from perturbations in the standard cosmic fluid (e.g., independent of the dark matter density or local clock speed). In paper III, we will discuss how distortion anisotropies from decaying or annihilating particles will manifest, and then provide simple constraints on these cases.

Thirdly, the effect of distortions (and thermalisation effects) on the evolution of the standard perturbation variables is negligible unless non-standard sources of temperature perturbations are considered. This means that in many scenarios with distortion anisotropies, the transfer functions of the distortion variables will behave like a driven oscillator, following closely the corresponding temperature variables. This approximation is possible unless we look at the generation and thermalisation of large distortions during the early phases ($z \gtrsim 5 \times 10^6$), where momentarily one could still imagine average distortions $\mu_0^{(0)} \simeq 10^{-2}$ [59, 60] and hence efficient conversion into temperature perturbations of noticeable level. However, we leave a more detailed discussion of this regime to future work.

Moving away from the main time-dependent aspects, let us consider the scale-dependence of the distortion evolution. In the absence of anisotropic heating, distortion anisotropies are only generated once average distortions are present. This means that there will be a difference in the transfer functions depending on when during the cosmic evolution of the mode the injection occurs. Crucial moments are horizon crossing, diffusion damping and free-streaming, all known from the standard CMB temperature anisotropies [e.g., 33, 61].

If we pick a mode of wavenumber k, we can distinguish three main regimes:

i) the average distortion is generated after the mode has fully damped away by Silk-damping ($k \gg k_{\rm D}$, where $k_{\rm D} \simeq 4.0 \times 10^{-6}(1+z)^{3/2}\,{\rm Mpc}^{-1}$ is the Silk-damping scale [54, 62]). In this case, no distortion anisotropies are formed because fluctuations in the plasma are no longer present at the corresponding scale.

ii) the average distortion is generated when the mode is well within the horizon but at $k \lesssim k_{\rm D}$. In this regime, no sources of distortions from potential variations arise, and distortion anisotropies are generated solely by Doppler terms and perturbed thermalisation effects while at $z \gtrsim 10^4$.

iii) the average distortion is generated while the mode is still super-horizon. In this case, both Doppler and potential perturbations affect the distortion fluctuations. We will furthermore be sensitive to perturbed thermalisation effects when considering the evolution at $z \gtrsim 10^4$.

We will quantify these general expectations in paper III. We also note that once anisotropic heating is included, the statistical properties of the source of heat could be highly non-Gaussian in which case a more general Boltzmann system may have to be solved. In addition, direct sourcing of distortions in the regime $k \gg k_{\rm D}$ can still lead to significant distortions at very small scales, e.g., by phase transitions or non-linear baryonic physics.

3 Discussion and conclusions

This work provides the main formulation of a new photon Boltzmann hierarchy, Eq. (2.26) and Eq. (2.30), that allows us to compute the evolution of distortion anisotropies at first order in perturbation theory (Sect. 2). Distortion anisotropy sources from Doppler and potential terms as well as anisotropic heating and perturbed thermalisation are accounted for. We also account for the spectral evolution of the distortion based on an approximate ODE treatment for the thermalisation Green's function, which captures most aspects of the full calculation using an extended spectral basis to describe the residual distortion evolution (see paper I). We furthermore demonstrate that the line-of-sight approach can be generalized to simplify the computation of the CMB signal power spectra

(Sect. 2.8). We briefly explain how the computation can be simplified by converting to a spectral basis that is optimised for the experiment using a basis rotation as explained in paper I (Sect. 2.8.1)

Overall, this paper is the second step in a series of works discussing the production and evolution of SD anisotropies generated by various physical mechanisms and how these might be constrained with future CMB spectrometers and imagers, enabling more realistic SD anisotropy forecasts over a wide range of physics. The new formulation furthermore is a first step towards a more general and precise treatment of CMB temperature anisotropies at second order in perturbation theory. In the standard approach [35, 38, 39], an 'instantaneous thermalisation' approximation is applied which ensures full energy conservation for the photons but without allowing a rigorous separation of distortion and temperature terms. With our new ODE representation of the distortion Green's function, one should be able to overcome this problem, even if additional generalisations will be needed.

We highlight that Doppler boosting and potential driving distortion source terms were omitted in most previous discussions of primordial SD anisotropies, although the importance of these terms was recognized earlier [24]. These terms are indeed small with respect to the standard temperature perturbation (hence not affecting the their evolution notably); however, for the SD anisotropies they provide the leading order source terms once average distortions are present. Since SDs can be spectrally distinguished, these terms remain relevant, leading to the independent distortion parameter hierarchies we presented here. The resultant distortion anisotropies are expected to be significant as long as the average distortion is at the level that *COBE/FIRAS* allows. A more quantitative discussion will be given in paper III, and subsequent works. Importantly, in paper III we will demonstrate that this effect allows one to derive limits on the average heating rate using existing and future CMB anisotropy measurements.

Finally, the generalised Boltzmann system can still be improved. We did not consider the effect of polarised distortions nor thermalisation/Compton scattering in the anisotropies, which could further modify the result. We furthermore neglected kinematic corrections as well as small non-linear Comptonisation terms and Comptonisation effects in the anisotropies. A brute force treatment of the problem for single modes, even if computationally demanding, might reveal additional augmentations to the problem that may require extra attention. It would also be extremely important to formulate the problem in alternative gauges/frames, to check the consistency of the equations and also explore possibilities for simplifying the computations, finding analytic approximations, and clarifying the physical interpretation of the signals. We leave all these improvements of our method to the future; however, our main results should not be affected. We therefore conclude that one of the main steps towards quasi-exact computations of anisotropic SDs in the perturbed Universe is taken.

Acknowledgments

We cordially thank the referee for their useful comments and significant effort in trying to help make the manuscript more digestible. We also thank Eiichiro Komatsu, Aditya Rotti and Rashid Sunyaev for stimulating discussion. In addition we thank Antony Lewis for his questions about apparent super-horizon evolution of distortion modes, which encouraged us to carry out a more complete derivation of the equations. We furthermore thank Nicola Bartolo, Daniele Bertacca, Colin Hill, Rishi Khatri, Sabino Matarrese, Atsuhisa Ota, Enrico Pajer and Nils Schöneberg for comments on the manuscript. This work was supported by the ERC Consolidator Grant *CMBSPEC* (No. 725456). TK was also supported by STFC grant ST/T506291/1. JC was furthermore supported by the Royal Society as a Royal Society University Research Fellow at the University of Manchester, UK (No. URF/R/191023). AR acknowledges support by the project "Combining Cosmic Microwave Background and Large Scale Structure data: an Integrated Approach for Addressing Fundamental Questions in Cosmology", funded by the MIUR Progetti di Ricerca di Rilevante Interesse Nazionale (PRIN) Bando 2017 - grant 2017YJYZAH.

References

[1] C. L. Bennett, A. J. Banday, K. M. Gorski, et al., *Four-Year COBE DMR Cosmic Microwave Background Observations: Maps and Basic Results*, *The Astrophysical Journal Letters* **464** (June, 1996) L1, [`astro-ph/9`].

[2] C. L. Bennett, M. Halpern, G. Hinshaw, et al., *First-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Preliminary Maps and Basic Results*, *The Astrophysical Journal Supplement* **148** (Sept., 2003) 1–27.

[3] R. Hlozek, J. Dunkley, G. Addison, et al., *The Atacama Cosmology Telescope: a measurement of the primordial power spectrum*, *ArXiv:1105.4887* (May, 2011) [`arXiv:1105.4887`].

[4] R. Keisler, C. L. Reichardt, K. A. Aird, et al., *A Measurement of the Damping Tail of the Cosmic Microwave Background Power Spectrum with the South Pole Telescope*, *The Astrophysical Journal* **743** (Dec., 2011) 28, [`arXiv:1105.3182`].

[5] Planck Collaboration, P. A. R. Ade, N. Aghanim, et al., *Planck 2015 results. XIII. Cosmological parameters*, *Astronomy & Astrophysics* **594** (Sept., 2016) A13, [`arXiv:1502.01589`].

[6] K. N. Abazajian, P. Adshead, Z. Ahmed, et al., *CMB-S4 Science Book, First Edition*, *ArXiv:1610.02743* (Oct., 2016) [`arXiv:1610.02743`].

[7] J. Delabrouille, P. de Bernardis, F. R. Bouchet, et al., *Exploring cosmic origins with CORE: Survey requirements and mission design*, *Journal of Cosmology and Astroparticle Physics* **4** (Apr., 2018) 014, [`arXiv:1706.04516`].

[8] S. Hanany, M. Alvarez, E. Artis, et al., *PICO: Probe of Inflation and Cosmic Origins*, *arXiv e-prints* (Feb., 2019) arXiv:1902.10541, [`arXiv:1902.10541`].

[9] J. Delabrouille, M. H. Abitbol, N. Aghanim, et al., *Microwave spectro-polarimetry of matter and radiation across space and time*, *Experimental Astronomy* **51** (June, 2021) 1471–1514.

[10] J. Chluba, A. Kogut, S. P. Patil, et al., *Spectral Distortions of the CMB as a Probe of Inflation, Recombination, Structure Formation and Particle Physics*, *BAAS* **51** (May, 2019) 184, [`arXiv:1903.04218`].

[11] J. Chluba, M. H. Abitbol, N. Aghanim, et al., *New horizons in cosmology with spectral distortions of the cosmic microwave background*, *Experimental Astronomy* **51** (June, 2021) 1515–1554, [`arXiv:1909.01593`].

[12] J. Chluba, T. Kite, and A. Ravenni, *Spectro-spatial evolution of the CMB I: discretisation of the thermalisation Green's function*, *arXiv e-prints* (Oct., 2022) arXiv:2210.09327, [`arXiv:2210.09327`].

[13] **LiteBIRD** Collaboration, M. Hazumi et al., *LiteBIRD: JAXA's new strategic L-class mission for all-sky surveys of cosmic microwave background polarization*, *Proc. SPIE Int. Soc. Opt. Eng.* **11443** (2020) 114432F, [`arXiv:2101.12449`].

[14] **NASA PICO** Collaboration, S. Hanany et al., *PICO: Probe of Inflation and Cosmic Origins*, `arXiv:1902.10541`.

[15] R. A. Sunyaev and Y. B. Zeldovich, *The interaction of matter and radiation in the hot model of the Universe, II*, *Astrophysics and Space Science* **7** (Apr., 1970) 20–30.

[16] L. Danese and G. de Zotti, *Double Compton process and the spectrum of the microwave background*, *Astronomy & Astrophysics* **107** (Mar., 1982) 39–42.

[17] C. Burigana, L. Danese, and G. de Zotti, *Formation and evolution of early distortions of the microwave background spectrum - A numerical study*, *Astronomy & Astrophysics* **246** (June, 1991) 49–58.

[18] W. Hu and J. Silk, *Thermalization and spectral distortions of the cosmic background radiation*, *Physical Review D* **48** (July, 1993) 485–502.

[19] J. Chluba, *Refined approximations for the distortion visibility function and μ-type spectral distortions*, *Monthly Notices of the Royal Astronomical Society* **440** (Apr., 2014) 2544–2563, [`arXiv:1312.6030`].

[20] R. Khatri and R. A. Sunyaev, *Beyond y and μ: the shape of the CMB spectral distortions in the intermediate epoch, 1.5 x 10^4lesssimzlesssim2 x 10^5*, *Journal of Cosmology and Astroparticle Physics* **9** (Sept., 2012) 16, [`arXiv:1207.6654`].

[21] A. Ota, *CMB spectral distortions as solutions to the Boltzmann equations*, *Journal of Cosmology and Astroparticle Physics* **2017** (Jan., 2017) 037, [`arXiv:1611.08058`].

[22] T. Haga, K. Inomata, A. Ota, and A. Ravenni, *Exploring compensated isocurvature perturbations with CMB spectral distortion anisotropies*, *Journal of Cosmology and Astroparticle Physics* **2018** (Aug., 2018) 036, [`arXiv:1805.08773`].

[23] T. Kite, A. Ravenni, and J. Chluba, *Spectro-spatial evolution of the CMB III: transfer functions, power spectra and Fisher forecasts*, *arXiv e-prints* (Dec., 2022) arXiv:2212.02817, [`arXiv:2212.02817`].

[24] J. Chluba, R. Khatri, and R. A. Sunyaev, *CMB at 2 x 2 order: the dissipation of primordial acoustic waves and the observable part of the associated energy release*, *Monthly Notices of the Royal Astronomical Society* **425** (Sept., 2012) 1129–1169, [`arXiv:1202.0057`].

[25] J. Chluba and R. M. Thomas, *Towards a complete treatment of the cosmological recombination problem*, *Monthly Notices of the Royal Astronomical Society* **412** (Apr., 2011) 748–764, [`arXiv:1010.3631`].

[26] J. Chluba, A. Ravenni, and B. Bolliet, *Improved calculations of electron-ion bremsstrahlung Gaunt factors for astrophysical applications*, *Monthly Notices of the Royal Astronomical Society* **492** (Feb., 2020) 177–194, [`arXiv:1911.08861`].

[27] A. Ravenni and J. Chluba, *The double Compton process in astrophysical plasmas*, *Journal of Cosmology and Astroparticle Physics* **2020** (Oct., 2020) 025, [`arXiv:2005.06941`].

[28] J. Chluba and R. A. Sunyaev, *The evolution of CMB spectral distortions in the early Universe*, *Monthly Notices of the Royal Astronomical Society* **419** (Jan., 2012) 1294–1314, [`arXiv:1109.6552`].

[29] J. Chluba, *Green's function of the cosmological thermalization problem - II. Effect of photon injection and constraints*, *Monthly Notices of the Royal Astronomical Society* **454** (Dec., 2015) 4182–4196, [`arXiv:1506.06582`].

[30] B. Bolliet, J. Chluba, and R. Battye, *Spectral distortion constraints on photon injection from low-mass decaying particles*, *arXiv e-prints* (Dec., 2020) arXiv:2012.07292, [`arXiv:2012.07292`].

[31] C.-P. Ma and E. Bertschinger, *Cosmological Perturbation Theory in the Synchronous and Conformal Newtonian Gauges*, *The Astrophysical Journal* **455** (Dec., 1995) 7–+, [`astro-ph/9`].

[32] W. Hu and M. White, *Acoustic Signatures in the Cosmic Microwave Background*, *The Astrophysical Journal* **471** (Nov., 1996) 30–+, [`astro-ph/9`].

[33] S. Dodelson, *Modern cosmology*. Academic Press, 2003.

[34] W. Hu, D. Scott, and J. Silk, *Reionization and cosmic microwave background distortions: A complete treatment of second-order Compton scattering*, *Physical Review D* **49** (Jan., 1994) 648–670, [`astro-ph/9`].

[35] L. Senatore, S. Tassev, and M. Zaldarriaga, *Non-gaussianities from perturbing recombination*, *Journal of Cosmology and Astroparticle Physics* **9** (Sept., 2009) 38, [`arXiv:0812.3658`].

[36] C. Pitrou, *Radiative transport of relativistic species in cosmology*, *Astroparticle Physics* **125** (Feb., 2021) 102494, [`arXiv:1902.09456`].

[37] W. Hu and M. White, *CMB anisotropies: Total angular momentum method*, *Physical Review D* **56** (July, 1997) 596–615, [`astro-ph/9702170`].

[38] N. Bartolo, S. Matarrese, and A. Riotto, *Cosmic microwave background anisotropies at second order: I*, *Journal of Cosmology and Astroparticle Physics* **6** (June, 2006) 24, [`astro-ph/0604416`].

[39] C. Pitrou, *The radiative transfer at second order: a full treatment of the Boltzmann equation with polarization*, *Classical and Quantum Gravity* **26** (Mar., 2009) 065006, [arXiv:0809.3036].

[40] C. Pitrou and A. Stebbins, *Parameterization of temperature and spectral distortions in future CMB experiments*, *arXiv e-prints* (Feb., 2014) arXiv:1402.0968, [arXiv:1402.0968].

[41] W. Hu and N. Sugiyama, *Small-Scale Cosmological Perturbations: an Analytic Approach*, *The Astrophysical Journal* **471** (Nov., 1996) 542, [astro-ph/9].

[42] J. Chluba, *Tests of the CMB temperature-redshift relation, CMB spectral distortions and why adiabatic photon production is hard*, *Monthly Notices of the Royal Astronomical Society* **443** (Sept., 2014) 1881–1888, [arXiv:1405.1277].

[43] J. A. S. Lima, *Thermodynamics of decaying vacuum cosmologies*, *Physical Review D* **54** (Aug., 1996) 2571–2577, [gr-qc/9605055].

[44] S. R. G. Trevisani and J. A. S. Lima, *Gravitational Matter Creation, Multi-fluid Cosmology and Kinetic Theory*, *arXiv e-prints* (Mar., 2023) arXiv:2303.01974, [arXiv:2303.01974].

[45] J. Chluba and R. A. Sunyaev, *Superposition of blackbodies and the dipole anisotropy: A possibility to calibrate CMB experiments*, *Astronomy & Astrophysics* **424** (Sept., 2004) 389–408, [astro-ph/0].

[46] A. Stebbins, *CMB Spectral Distortions from the Scattering of Temperature Anisotropies*, *astro-ph/0703541* (Mar., 2007) [astro-ph/0703541].

[47] R. P. Bauman, R. L. Porter, G. J. Ferland, and K. B. MacAdam, *J-Resolved He I Emission Predictions in the Low-Density Limit*, *The Astrophysical Journal* **628** (July, 2005) 541–554, [astro-ph/0504083].

[48] Y. Watanabe and E. Komatsu, *Improved calculation of the primordial gravitational wave spectrum in the standard model*, *Physical Review D* **73** (June, 2006) 123515, [astro-ph/0604176].

[49] T. Kite, J. Chluba, A. Ravenni, and S. P. Patil, *Clarifying transfer function approximations for the large-scale gravitational wave background in* Λ*CDM*, *Monthly Notices of the Royal Astronomical Society* **509** (Jan., 2022) 1366–1376, [arXiv:2107.13351].

[50] B. Novosyadlyj, *Perturbations of ionization fractions at the cosmological recombination epoch*, *Monthly Notices of the Royal Astronomical Society* **370** (Aug., 2006) 1771–1782.

[51] R. Khatri and B. D. Wandelt, *Crinkles in the last scattering surface: Non-Gaussianity from inhomogeneous recombination*, *Physical Review D* **79** (Jan., 2009) 023501, [arXiv:0810.4370].

[52] A. Lewis, A. Challinor, and A. Lasenby, *Efficient Computation of Cosmic Microwave Background Anisotropies in Closed Friedmann-Robertson-Walker Models*, *The Astrophysical Journal* **538** (Aug., 2000) 473–476, [astro-ph/9911177].

[53] J. Lesgourgues, *The Cosmic Linear Anisotropy Solving System (CLASS) I: Overview*, *ArXiv:1104.2932* (Apr., 2011) [arXiv:1104.2932].

[54] N. Kaiser, *Small-angle anisotropy of the microwave background radiation in the adiabatic theory*, *Monthly Notices of the Royal Astronomical Society* **202** (Mar., 1983) 1169–1180.

[55] S. Weinberg, *Entropy Generation and the Survival of Protogalaxies in an Expanding Universe*, *The Astrophysical Journal* **168** (Sept., 1971) 175.

[56] S. Renaux-Petel, C. Fidler, C. Pitrou, and G. W. Pettinari, *Spectral distortions in the cosmic microwave background polarization*, *Journal of Cosmology and Astroparticle Physics* **2014** (Mar., 2014) 033, [arXiv:1312.4448].

[57] U. Seljak and M. Zaldarriaga, *A Line-of-Sight Integration Approach to Cosmic Microwave Background Anisotropies*, *The Astrophysical Journal* **469** (Oct., 1996) 437–+, [astro-ph/9].

[58] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions.* Dover Press, 1972.

[59] J. Chluba, A. Ravenni, and S. K. Acharya, *Thermalization of large energy release in the early Universe*, *Monthly Notices of the Royal Astronomical Society* **498** (Oct., 2020) 959–980, [arXiv:2005.11325].

[60] S. K. Acharya and J. Chluba, *CMB spectral distortions from continuous large energy release*, *arXiv e-prints* (Dec., 2021) arXiv:2112.06699, [arXiv:2112.06699].

[61] W. Hu and N. Sugiyama, *Anisotropies in the cosmic microwave background: an analytic approach*, *The Astrophysical Journal* **444** (May, 1995) 489–506, [astro-ph/9].

[62] J. Silk, *Cosmic Black-Body Radiation and Galaxy Formation*, *The Astrophysical Journal* **151** (Feb., 1968) 459–+.

[63] G. W. Pettinari, *The intrinsic bispectrum of the cosmic microwave background*. PhD thesis, University of Portsmouth, UK, Jan., 2013.

[64] J. Chluba, L. Dai, and M. Kamionkowski, *Multiple scattering Sunyaev-Zeldovich signal - I. Lowest order effect*, *Monthly Notices of the Royal Astronomical Society* **437** (Jan., 2014) 67–76, [arXiv:1308.5969].

[65] A. Kompaneets, *The establishment of thermal equilibrium between photons and electrons*, *Sov.Phys. JETP* **31** (1956) 876.

[66] A. P. Lightman, *Double Compton emission in radiation dominated thermal plasmas*, *The Astrophysical Journal* **244** (Mar., 1981) 392–405.

[67] J. Chluba, J. A. Rubiño-Martín, and R. A. Sunyaev, *Cosmological hydrogen recombination: populations of the high-level substates*, *Monthly Notices of the Royal Astronomical Society* **374** (Feb., 2007) 1310–1320, [astro-ph/0608242].

[68] Y. B. Zeldovich and E. V. Levich, *Stationary state of electrons in a non-equilibrium radiation field.*, *SJETPL* **11** (1970) 35–38.

[69] S. Y. Sazonov and R. A. Sunyaev, *Gas Heating Inside Radio Sources to Mildly Relativistic Temperatures via Induced Compton Scattering*, *Astronomy Letters* **27** (Aug., 2001) 481–492, [astro-ph/0].

[70] S. Seager, D. D. Sasselov, and D. Scott, *How Exactly Did the Universe Become Neutral?*, *The Astrophysical Journal Supplement* **128** (June, 2000) 407–430, [astro-ph/9912182].

[71] J. Chluba, *Spectral Distortions of the Cosmic Microwave Background*. PhD thesis, LMU München, Mar., 2005.

[72] R. Khatri, R. A. Sunyaev, and J. Chluba, *Does Bose-Einstein condensation of CMB photons cancel μ distortions created by dissipation of sound waves in the early Universe?*, *Astronomy & Astrophysics* **540** (Apr., 2012) A124, [arXiv:1110.0475].

[73] R. Khatri and R. A. Sunyaev, *Creation of the CMB spectrum: precise analytic solutions for the blackbody photosphere*, *Journal of Cosmology and Astroparticle Physics* **6** (June, 2012) 38, [arXiv:1203.2601].

A Derivation of the photon collision term at first order

In this section, we obtain the required terms of the collision term at first order in perturbations. For this, we start from Appendix C3 through C7 of [24] and also consider some modifications due to the transformation to the local inertial frame, adapting the treatments of [35, 63]. Additional details about relativistic radiative transfer can be found in [36]. The goal is to include the thermalisation effects and energy exchange for the local monopole and also account for all distortion sources that arise at background level. In the derivation below, we neglect polarisation terms and non-linear corrections to the distortion evolution, e.g., coming from the Kompaneets term due to stimulated recoil. This means we drop terms $O(\Delta n_0^2)$, linearising the scattering problem; however, we explain why this seems justified in our treatment of the problem, and also briefly outline how one might be able to account for these effects in the future.

A.1 Thomson terms

Starting from the standard physics of CMB temperature anisotropies, the Thomson terms and first order Doppler boosts carry over trivially, leading to [e.g., 38]

$$\dot{\tau}^{-1}\, C^{(1)}[n]\Big|_{\rm T} = n_0^{(1)} + \frac{1}{10}\, n_2^{(1)} - n^{(1)} + \beta^{(1)}\chi\, \hat{O}_x n^{(0)} \tag{A.1}$$

without further ado. Here, we introduced the direction cosine $\chi = \hat{\boldsymbol{\gamma}} \cdot \hat{\boldsymbol{\beta}}$ between the velocity and photon direction. We also defined $n_\ell(t, x, \boldsymbol{r}, \hat{\boldsymbol{\gamma}}) = \sum_m n_{\ell m}(t, x, \boldsymbol{r}) Y_{\ell m}(\hat{\boldsymbol{\gamma}})$ using the spherical harmonic coefficients of the photon occupation number, $n_{\ell m}(t, x, \boldsymbol{r})$. It is important that, in contrast to the usual treatment, we now include the average distortion in $\hat{O}_x n^{(0)} = G + \hat{O}_x \Delta n^{(0)}$, as this leads to distortion anisotropies, which would not arise otherwise. This is the leading order source term at the level $\propto \dot{\tau}\beta^{(1)}\, \hat{O}_x \Delta n_0^{(0)}$, which becomes noticeable once the mode enters the horizon. Without average distortions, this term vanished and consequently only temperature fluctuations are sourced.

In Eq. (A.1), we neglected small Klein-Nishina corrections $\propto \theta_{\rm e} = k_{\rm B} T_{\rm e}/m_{\rm e} c^2$, as these do not change the spectral shape of the photon field but merely modify the Thomson scattering rates of the dipole, quadrupole and octupole [24, 64]. Some of these corrections have been considered in [22].

A.2 Kompaneets terms

Let us next consider the Comptonisation terms from the scattering of electrons and photon with non-zero energy exchange, which leads to spectral evolution. We shall include these effects only for the local monopole as Thomson terms are not suppressed for $\ell > 0$ and hence dominate the evolution there. The standard Kompaneets equation is [34, 65]

$$\begin{aligned}\dot{\tau}^{-1}\, C[n]\Big|_{\rm K} &= \theta_{\rm e} \hat{\mathcal{D}}_x n_0 + \theta_z \hat{\mathcal{D}}^*_x n_0 (1 + n_0) = \Delta\theta_{\rm e} Y + \theta_{\rm e} \hat{\mathcal{D}}_x \Delta n_0 + \theta_z \hat{\mathcal{D}}^*_x A\, \Delta n_0 + \theta_z \hat{\mathcal{D}}^*_x \Delta n_0^2 \\ &= \Delta\theta_{\rm e} \left[Y + \hat{\mathcal{D}}_x \Delta n_0\right] + \theta_z \hat{\mathcal{K}}_x \Delta n_0 + \theta_z \hat{\mathcal{D}}^*_x \Delta n_0^2.\end{aligned} \tag{A.2}$$

where we used $\hat{\mathcal{D}}_x = x^{-2}\partial_x x^4 \partial_x$, $\hat{\mathcal{D}}^*_x = x^{-2}\partial_x x^4$, $\hat{\mathcal{K}}_x = \hat{\mathcal{D}}_x + \hat{\mathcal{D}}^*_x A$ and $A = 1 + 2n_{\rm bb}$. From the first group of terms we see that the local difference between the electron and photon temperature is the main source of distortions, leading to injection of Y with a source correction from $\hat{\mathcal{D}}_x \Delta n_0$. We have not yet included the effect of transforming to the local inertial frame, which causes an overall factor of $(1 + \Psi)$ that usually only becomes important at second order in perturbation theory [35, 63]. However, here we are dealing with a non-vanishing zeroth order term, which then required inclusion of this factor already at first order in perturbation once non-zero distortions are present.

Adding this factor and collecting terms, we then have the Compton collision terms at zeroth and first order in perturbations,

$$\left.\frac{C^{(0)}[n]}{\dot{\tau}}\right|_{\rm CS} = \Delta\theta_{\rm e}^{(0)} Y + \theta_z \hat{\mathcal{K}}_x \Delta n_0^{(0)} + \Delta\theta_{\rm e}^{(0)} \hat{\mathcal{D}}_x \Delta n_0^{(0)} + \theta_z \hat{\mathcal{D}}_x^{*} (\Delta n_0^{(0)})^2 \approx \Delta\theta_{\rm e}^{(0)} Y + \theta_z \hat{\mathcal{K}}_x \Delta n_0^{(0)} \tag{A.3}$$

$$\begin{aligned}\left.\frac{C^{(1)}[n]}{\dot{\tau}}\right|_{\rm CS} &= \Delta\theta_{\rm e}^{(1)} Y + \theta_z \hat{\mathcal{K}}_x \Delta n_0^{(1)} + \Delta\theta_{\rm e}^{(1)} \hat{\mathcal{D}}_x \Delta n_0^{(0)} + \Delta\theta_{\rm e}^{(0)} \hat{\mathcal{D}}_x \Delta n_0^{(1)} + 2\theta_z \hat{\mathcal{D}}_x^{*} \Delta n_0^{(0)} \Delta n_0^{(1)} \\ &\qquad + \left[\frac{\dot{\tau}^{(1)}}{\dot{\tau}} + \Psi^{(1)}\right] \left(\Delta\theta_{\rm e}^{(0)}\, Y + \theta_z\, \hat{\mathcal{K}}_x\, \Delta n_0^{(0)}\right) \\ &\approx \Delta\theta_{\rm e}^{(1)} Y + \theta_z \hat{\mathcal{K}}_x \Delta n_0^{(1)} + \left[\frac{\dot{\tau}^{(1)}}{\dot{\tau}} + \Psi^{(1)}\right] \left.\frac{C^{(0)}[n]}{\dot{\tau}}\right|_{\rm CS} + \Delta\theta_{\rm e}^{(1)} \hat{\mathcal{D}}_x \Delta n_0^{(0)} + 4\Theta_0^{(1)} \Delta\theta_{\rm e}^{(0)}\, Y_1. \end{aligned} \tag{A.4}$$

At zeroth order, we neglect the small correction $\propto \Delta\theta_{\rm e}^{(0)} \hat{\mathcal{D}}_x \Delta n_0^{(0)} + \theta_z \hat{\mathcal{D}}_x^{*} (\Delta n_0^{(0)})^2$, which does not modify the leading order picture significantly, being second order in the average distortion. Also in the standard thermalisation computation with `CosmoTherm` this approximation works to extremely high precision unless large distortions are encountered [59, 60].

At first order, the leading source term is again due to differences in the local electron temperature with respect to T_z. However, since Compton equilibrium is reached quickly, $\Delta\theta_{\rm e}^{(1)}$ will be comparable to that of the local photon temperature perturbation with corrections from the local distortion and heating rate (see Appendix C.2). This source term therefore mainly captures the modulation of the local thermal equilibrium by the variation of the ambient blackbody temperature.

The term $\Delta\theta_{\rm e}^{(0)} \hat{\mathcal{D}}_x \Delta n_0^{(1)}$ leads to a correction to the Comptonisation time-scale[7] of $\Delta n_0^{(1)}$, with the dominant term given by $\theta_z \hat{\mathcal{K}}_x \Delta n_0^{(1)}$. Since $\Delta\theta_{\rm e}^{(0)} \simeq O(\Delta n_0^{(0)})$, here we can use $\Delta n_0^{(1)} \approx \Theta_0^{(1)} G$, such that we obtain the source

$$\hat{\mathcal{D}}_x \Delta n_0^{(1)} \approx \Theta_0^{(1)}\, \hat{\mathcal{D}}_x \hat{\mathcal{O}}_x n_{\rm bb} \equiv \Theta_0^{(1)}\, \hat{\mathcal{O}}_x \hat{\mathcal{D}}_x n_{\rm bb} = 4\Theta_0^{(1)}\, Y_1, \tag{A.5}$$

as used in the last step of Eq. (A.4). We can also rewrite the term $\hat{\mathcal{D}}_x \Delta n_0^{(0)} = (\hat{\mathcal{O}}_x - 3)\hat{\mathcal{O}}_x \Delta n_0^{(0)}$ with the method described in the main text to simplify the correction $\Delta\theta_{\rm e}^{(1)} \hat{\mathcal{D}}_x \Delta n_0^{(0)}$, which is related to perturbed scattering terms.

The term $\propto 2\theta_z \hat{\mathcal{D}}_x^{*} \Delta n_0^{(0)} \Delta n_0^{(1)}$ modifies the stimulated Compton scattering rate of $\Delta n_0^{(1)}$. This is dominated by the blackbody part, $\propto \hat{\mathcal{D}}_x^{*} A\, \Delta n_0^{(1)}$, within the Kompaneets operator, so that we neglect this correction. It is clear that this term in principle is of similar order as terms that we indeed keep; however, since it merely modifies the exact timing of leading order scattering terms, we believe that its omission does not change the main conclusions significantly. A full numerical solution will be required to check the error this simplification introduces. For this, we can in principle start by mapping this term back onto the spectral basis just like for the Kompaneets operator, $\hat{\mathcal{K}}_x$. This will yield a representation in matrix form, creating a rotation of the spectral vector like for $\hat{\mathcal{K}}_x$; however, we leave an exploration of this effect to the future.

The term $\propto \dot{\tau}^{(1)}/\dot{\tau}$ are also due to scattering time-scale modulations but this time from perturbations in the electron density, which we include as $\dot{\tau}^{(1)}/\dot{\tau} = N_{\rm e}^{(1)}/N_{\rm e} \approx \delta_{\rm b}^{(1)}$ using the baryon density perturbation, $\delta_{\rm b}^{(1)}$. We shall neglect perturbed recombination effects here [e.g., 35, 50, 51]. These should never become important in the regimes we are interested in, when θ_z is not too small already. The terms $\propto \Psi^{(1)}$ account for the transformation to the local inertial frame [35], which were omitted in [24]. These terms lead to perturbed scattering effects once an average distortion is present.

---

[7] More precisely the Doppler broadening and boosting term, $\propto \hat{\mathcal{D}}_x$.

To close the discussion of all the effects from Compton scattering, we mention that additional velocity corrections appear that are related to dipole scattering and kinematic effects. However, to simplify the problem we neglect these terms. From [24], it is clear that these terms can only be relevant at high redshifts, where significant energy exchange occurs, as we briefly discuss now.

A.2.1 Kinematic corrections to the Kompaneets term

As already mentioned in the main text, we do not include any energy exchange effects mediated by Compton scattering on the anisotropies in the spectrum. This means that in the rest frame of the moving thermal electron distribution one has

$$\left.\frac{\partial n'(x',\hat{\gamma}')}{\partial y'_z}\right|_{\rm K} \approx \frac{\theta_{\rm e}}{\theta_z}\,\hat{\mathcal{D}}_{x'}\,n'_0(x') + \hat{\mathcal{D}}^*_{x'}\,n'_0(x')[1+n'_0(x')], \tag{A.6}$$

where the primes on variables denote that they are evaluated in the moving frame. After a boost, the average occupation number in the moving frame is $n'_0(x') \approx \int n[x(x',\beta,\chi'),\chi(\beta,\chi')]\,\frac{{\rm d}\chi'}{2} \approx n_0(x')$, where $x(x',\beta,\chi') \approx x'(1+\beta\chi')$ and $\chi(\beta,\chi') = (\chi'+\beta)/(1+\beta\chi')$. Here we neglected possible corrections $\mathcal{O}(\beta n_1)$ from aberration effects on the restframe dipole spectrum to the monopole in the moving frame. Using the invariance of $\hat{\mathcal{D}}_{x'} \equiv \hat{\mathcal{D}}_x$ and $\hat{\mathcal{D}}^*_{x'} \approx (1-\beta\chi)\hat{\mathcal{D}}^*_x$ together with the transformation of the scattering y-parameter we then have

$$\begin{aligned}
\left.\frac{\partial n(x,\hat{\gamma})}{\partial y_z}\right|_{\rm K} &= (1-\beta\chi)\left.\frac{\partial n'(x',\hat{\gamma}')}{\partial y'_z}\right|_{\rm K} \approx (1-\beta\chi)\left\{\frac{\theta_{\rm e}}{\theta_z}\,\hat{\mathcal{D}}_x\,n_0(x') + (1-\beta\chi)\,\hat{\mathcal{D}}^*_x\,n_0(x')[1+n_0(x')]\right\}\\
&\approx \frac{\theta_{\rm e}}{\theta_z}\,\hat{\mathcal{D}}_x\,n_0(x) + \hat{\mathcal{D}}^*_x\,n_0(x)[1+n_0(x)] - \beta\chi\left\{\frac{\theta_{\rm e}}{\theta_z}\,\hat{\mathcal{D}}_x\,n_0(x) + \hat{\mathcal{D}}^*_x\,n_0(x)[1+n_0(x)]\right\}\\
&\qquad - \beta\chi\,\hat{\mathcal{D}}^*_x\,n_0(x)[1+n_0(x)] + \beta\chi\left\{\frac{\theta_{\rm e}}{\theta_z}\,\hat{\mathcal{D}}_x + \hat{\mathcal{D}}^*_x[1+2\,n_0(x)]\right\}\mathcal{T}(x)
\end{aligned} \tag{A.7}$$

where in the last step we used $x' \approx x(1-\beta\chi)$ and $n_0(x') \approx n_0(x) + \beta\chi\,\mathcal{T}(x)$ with $\mathcal{T}(x) = \hat{O}_x n_0(x)$. This result is consistent with Eq. (C37) in [24] after omitting dipole scattering effects. Assuming small departures from the average blackbody spectrum, $\Delta n_0 \ll 1$, we can insert $n_0 = n_{\rm bb} + \Delta n_0$ and then linearize in Δn_0, which with $\mathcal{T} = G + \hat{O}_x \Delta n_0$ yields:

$$\begin{aligned}
\left.\frac{\partial \Delta n(x,\hat{\gamma})}{\partial y_z}\right|_{\rm K} &\approx \frac{\Delta T_{\rm e}}{T_z}\,Y(x) + \hat{\mathcal{K}}_x\,\Delta n_0 - 2\beta\chi\left\{\frac{\Delta T_{\rm e}}{T_z}\,Y(x) + \hat{\mathcal{K}}_x\,\Delta n_0\right\} + \frac{T_{\rm e}}{T_z}\,\beta\chi\left[Y(x) + \hat{\mathcal{D}}_x\,\Delta n_0\right]\\
&\qquad + \beta\chi\left\{\frac{\Delta T_{\rm e}}{T_z}\,\hat{\mathcal{D}}_x + \hat{\mathcal{K}}_x + 2\hat{\mathcal{D}}^*_x\,\Delta n_0\right\}G + \beta\chi\,\hat{\mathcal{K}}_x\hat{O}_x\Delta n_0 \\
&\approx \frac{\Delta T_{\rm e}}{T_z}\,Y + \hat{\mathcal{K}}_x\,\Delta n_0 - \beta\chi\left\{\frac{\Delta T_{\rm e}}{T_z}\,[Y - 4Y_1] + \left[\hat{\mathcal{K}}_x - \hat{\mathcal{K}}_x\hat{O}_x + \hat{\mathcal{D}}^*_x\,(A-2G)\right]\Delta n_0\right\}
\end{aligned} \tag{A.8}$$

The first two terms are just the standard Kompaneets terms, while the last group of terms accounts for first order velocity corrections. In computations, the terms in braces would be evaluated at background level and are of order $\beta\chi\,\theta_z$ per Thomson event. However, given that from Thomson terms we have a source $\simeq \beta\chi\,\mathcal{T}(x)$, these temperature correction terms are merely a small modification to the main source term (suppressed by a factor of $\theta_z \ll 1$). It is therefore well-justified to neglect these contribution unless high precision is needed, as already mentioned above. A more rigorous study of these effects is deemed important for obtaining a self-consistent (gauge-independent) formulation of the problem, in particular once scattering corrections from the anisotropies [24] are included.

A.3 Photon number changing processes

Aside from scattering, which conserves photon number, we also have to treat the conversion of distortions into pure temperature shifts. This thermalisation process is mediated by the combined action of Compton scattering and photon emission processes (DC and BR). Modelling the exact evolution of the photon field when DC and BR are included is difficult; however, it is known that the evolution of the high frequency spectrum ($\nu \gtrsim 1\,\mathrm{GHz}$) is not affected by these processes once the μ-era ends [15, 18, 19]. We can therefore obtain an approximate description that simply leads to a *redistribution of energy* between the μ parameter and local photon temperature. The net photon emission and absorption term takes the explicit form [18, 19, 28]

$$\left.\frac{\partial n_0}{\partial \tau}\right|_{\rm em} = \frac{\Lambda(x, \theta_{\rm e}, \theta_\gamma)\,{\rm e}^{-x\,\theta_z/\theta_{\rm e}}}{x^3}\left[1 - n_0\left({\rm e}^{x\,\theta_z/\theta_{\rm e}} - 1\right)\right] \tag{A.9}$$

in the local inertial frame. We introduced the photon emissivity, $\Lambda(x, \theta_{\rm e}, \theta_\gamma)$, which for DC scales as $\Lambda(x, \theta_{\rm e}, \theta_\gamma) \propto \theta_\gamma^2$ being driven by the high-frequency blackbody photons [16, 27, 66]. For convenience, we shall use the shorthand notation $\Lambda(x, \theta_z) \equiv \Lambda(x, \theta_z, \theta_z)$ below. We neglect kinematic correction to the emission process and also do not consider the energy exchange corrections from this term. These are expected to be negligible and also require a more careful study of the emission process. Some first steps have been outline in [24].

Photon production processes are in equilibrium if the CMB occupation number is given by a blackbody at the electron temperature. Thus, defining the distortion with respect to $n_{\rm bb}(x\,\theta_z/\theta_{\rm e})$ would simplify several aspects of the computation; however, the invariance of the spectrum under redshifting is no longer guaranteed, such that we will not use this alternative description.

Perturbing the equation and neglecting terms that are higher order in the average distortions, we can find the emission term at zeroth and first order in perturbations as

$$\left.\frac{\partial n_0^{(0)}}{\partial \tau}\right|_{\rm em} \approx -\frac{\Lambda(x, \theta_z)(1-{\rm e}^{-x})}{x^3}\,\Delta n_0^{(0)} + \frac{\Lambda(x, \theta_z)}{x^2}\,n_{\rm bb}\Theta_{\rm e}^{(0)} = -\frac{\Lambda(x, \theta_z)}{x^2}\,\frac{n_{\rm bb}}{G}\left[\Delta n_0^{(0)} - \Theta_{\rm e}^{(0)}\,G\right] \tag{A.10a}$$

$$\begin{aligned}\left.\frac{\partial n_0^{(1)}}{\partial \tau}\right|_{\rm em} &\approx -\frac{\Lambda(x, \theta_z)}{x^2}\,\frac{n_{\rm bb}}{G}\left[\Delta n_0^{(1)} - \Theta_{\rm e}^{(1)}\,G\right] + \Theta_0^{(1)}\frac{\Lambda(x, \theta_z)}{x^2}\left[\Delta n_0^{(0)}\,{\rm e}^{-x} + (G - A + 1)\,\Theta_{\rm e}^{(0)}\right] \\ &\quad - \left[\delta_{\rm b}^{(1)} + \Psi^{(1)}\right]\frac{\Lambda(x, \theta_z)}{x^2}\,\frac{n_{\rm bb}}{G}\left[\Delta n_0^{(0)} - \Theta_{\rm e}^{(0)}\,G\right] \\ &\quad\quad - \Theta_0^{(1)}\left[\left.\frac{\partial \ln\Lambda}{\partial \ln\theta_\gamma}\right|_{\theta_z} + \left.\frac{\partial \ln\Lambda}{\partial \ln\theta_{\rm e}}\right|_{\theta_z}\right]\frac{\Lambda(x, \theta_z)}{x^2}\,\frac{n_{\rm bb}}{G}\left[\Delta n_0^{(0)} - \Theta_{\rm e}^{(0)}\,G\right],\end{aligned} \tag{A.10b}$$

where we used $(1 - {\rm e}^{-x})/x = n_{\rm bb}/G$. For the first order equation, we can see that the temperature derivatives of the emission coefficient modify the thermalisation effect. In the DC era, one has $\partial \ln\Lambda/\partial \ln\theta_{\rm e} \approx 0$ and $\partial \ln\Lambda/\partial \ln\theta_\gamma \approx 2$, where the latter signifies that most of the DC emission is driven by the blackbody part of the CMB spectrum [27, 66, 67]. Corrections to the picture from BR will be neglected here, but can in principle be added using `BRpack` [26]. However, in the BR era, photon production is already nearly frozen, such that our approximations should not make a significant difference to the final distortion evolution.

B Evolution in various limits

B.1 Evolution equations in Thomson limit

Assuming that the average distortion is frozen, we can use the Ansatz $n^{(1)} \approx \Theta^{(1)}\, G(x) + \Sigma^{(1)}\, \hat{O}_x \Delta n^{(0)}$ in Eq. (2.6). We then have

$$\int x^2 n^{(1)}\,\mathrm{d}x \approx N_G\,\Theta^{(1)} + 3\Sigma^{(1)} \int x^2 \Delta n^{(0)}\,\mathrm{d}x \equiv N_G \Theta^{(1)} + 3N_{\Delta n^{(0)}}\,\Sigma^{(1)} \tag{B.1a}$$

$$\int x^3 n^{(1)}\,\mathrm{d}x \approx E_G\,\Theta^{(1)} + 4\Sigma^{(1)} \int x^3 \Delta n^{(0)}\,\mathrm{d}x = E_G \Theta^{(1)} + 4E_{\Delta n^{(0)}}\,\Sigma^{(1)}, \tag{B.1b}$$

where $N_f = \int x^2 f(x)\,\mathrm{d}x$ is the number integral of the spectral shape, $f(x)$. Unless the average CMB temperature was affected by the energy release process, we have $N_{\Delta n^{(0)}} \approx 0$ and otherwise $N_{\Delta n^{(0)}} \approx N_G\,\Theta_0^{(0)}$. Taking the number and energy density moments of the Boltzmann equation, Eq. (2.6), with

$$\mathcal{D}_t[X] = \frac{\partial X}{\partial t} + \frac{c\hat{\gamma}}{a}\cdot\nabla X \tag{B.2a}$$

$$\mathcal{L}[X,b] = \frac{\partial X}{\partial t} + \frac{c\hat{\gamma}}{a}\cdot\nabla X + b\left(\frac{\partial \Phi^{(1)}}{\partial t} + \frac{c\hat{\gamma}}{a}\cdot\nabla\Psi^{(1)}\right) \tag{B.2b}$$

$$C_{\rm T}[X,b] = \dot{\tau}\left[X_0 + \frac{1}{10}X_2 - X + b\beta^{(1)}\chi\right] \tag{B.2c}$$

and Eq. (B.1) this then yields

$$\mathcal{L}[\Theta^{(1)}] + 3\Theta_0^{(0)}\mathcal{L}[\Sigma^{(1)}] \approx C_{\rm T}[\Theta^{(1)}] + 3\Theta_0^{(0)} C_{\rm T}[\Sigma^{(1)}] \tag{B.3a}$$

$$\mathcal{L}[\Theta^{(1)}] + \frac{4E_{\Delta n^{(0)}}}{E_G}\mathcal{L}[\Sigma^{(1)}] \approx C_{\rm T}[\Theta^{(1)}] + \frac{4E_{\Delta n^{(0)}}}{E_G} C_{\rm T}[\Sigma^{(1)}], \tag{B.3b}$$

where we used the shorthand $\mathcal{L}[X] = \mathcal{L}[X,1]$ and $C_{\rm T}[X] = C_{\rm T}[X,1]$. By taking appropriate sums and differences of these equations we can then find

$$\mathcal{L}[\Theta^{(1)}] \approx C_{\rm T}[\Theta^{(1)}], \qquad \mathcal{L}[\Sigma^{(1)}] \approx C_{\rm T}[\Sigma^{(1)}], \tag{B.4}$$

as would have directly followed by comparing coefficients of the two types of spectra, G and $\hat{O}_x \Delta n^{(0)}$. As noted in the main text, $\Theta^{(1)}$ starts with initial conditions from inflation while $\Sigma^{(1)} = 0$ until the distortion is in place. The equations above do not allow solving the system while $\Delta n^{(0)}$ is evolving.

B.2 Changing the average temperature at second order in $\bar{\Theta}$

In Sect. 2.3.2, we showed that at first order in $\bar{\Theta}$, a photon source term $\propto G(x)$ leaves the spectrum invariant [confirm Eq. (2.11)]. Here we now ask what photon source do we need to obtain a blackbody up to second order in $\bar{\Theta} \ll 1$. Writing the Taylor series for a blackbody as $n \approx n_{\rm bb}(x) + [\bar{\Theta} + \bar{\Theta}^2]\, G(x) + \frac{1}{2}\bar{\Theta}^2\, Y(x)$ and taking the time derivative (denoted with dot at fixed x) we have

$$\dot{n} \approx G(x)[\dot{\bar{\Theta}} + 2\bar{\Theta}\dot{\bar{\Theta}}] + \bar{\Theta}\dot{\bar{\Theta}}\, Y(x) \equiv S_G\,[G(x) + f_Y\, Y(x)]. \tag{B.5}$$

If we demand that the average blackbody temperature should change like $\dot{\bar{\Theta}} = S_T$, then we can trivially write $S_G = (1 + 2\bar{\Theta})\, S_T$. Similarly, we then have $\bar{\Theta}\dot{\bar{\Theta}} \equiv S_G\, f_Y$, which implies $f_Y = \bar{\Theta}/[1 + 2\bar{\Theta}]$. The overall evolution equation is then given by $\dot{n} \approx S_T\,[(1 + 2\bar{\Theta})G(x) + \bar{\Theta}\, Y(x)]$, which leaves a blackbody spectrum with changing temperature unchanged at second order in $\bar{\Theta}$. Note that the 'blackbody source' is no longer independent of the solution for the blackbody temperature and requires a y-type contribution to shift the maximum of n towards higher frequencies. At second order, one thus cannot create a blackbody with a temperature-independent source.

However, we have an alternative way of describing the same situation. Defining the effective temperature and y-parameters as $\bar{\Theta}_{\rm eff} = \bar{\Theta}[1 + \bar{\Theta}]$ and $\bar{y}_{\rm eff} = \bar{\Theta}^2/2 \approx \bar{\Theta}^2_{\rm eff}/2$, we can similarly determine the required source functions for G and Y. For the source of G, we then simply need $S_G \equiv \dot{\bar{\Theta}}_{\rm eff} = (1 + 2\bar{\Theta})\, S_T = \sqrt{1 + 4\bar{\Theta}_{\rm eff}}\, S_T \approx (1 + 2\bar{\Theta}_{\rm eff})\, S_T$. Similarly, we have $S_Y = \dot{\bar{y}}_{\rm eff} = \bar{\Theta} S_T = (\sqrt{1 + 4\bar{\Theta}_{\rm eff}} - 1) S_T/2 \approx \bar{\Theta}_{\rm eff}\, S_T$. Since up to second order in $\bar{\Theta}$ we have $\bar{\Theta}^2/2 \approx \bar{\Theta}^2_{\rm eff}/2$, this also directly follows from $S_Y \approx \bar{\Theta}_{\rm eff}\, \dot{\bar{\Theta}}_{\rm eff} \approx \bar{\Theta}_{\rm eff}\, S_T$. We therefore find the same source term as above, but with $\bar{\Theta} \to \bar{\Theta}_{\rm eff}$. The solution for $\bar{\Theta}_{\rm eff}$ is then $\bar{\Theta}_{\rm eff} = [\exp\big(2 \int S_t \, {\rm d}t\big) - 1]/2 = [\exp\big(2\bar{\Theta}\big) - 1]/2 \approx \bar{\Theta}[1 + \bar{\Theta}]$, which as expected shows the equivalence of the two approaches.

C Physical approximations for the kinetic equation

C.1 Compton Electron temperature and Compton energy exchange

The energy exchange between electrons (and matter) and photons is controlled by the Compton scattering collision term. Smaller corrections due to BR and DC emission processes can be neglected. From Eq. (A.3), we can directly write the zeroth order energy exchange term as[8]

$$\Lambda^{(0)}_C = -\frac{8\pi h}{c^3}\left(\frac{k_{\rm B} T_z}{h}\right)^4 \int x^3 \dot{\tau} \left[\Delta\theta^{(0)}_{\rm e} Y + \theta_z \hat{\mathcal{K}}_x \Delta n^{(0)}_0\right] {\rm d}x \equiv \kappa \left[\Theta^{(0)}_{\rm C} - \Theta^{(0)}_{\rm e}\right] \tag{C.1a}$$

$$\kappa = 4\rho_z \dot{\tau}\, \theta_z, \qquad \Theta^{(0)}_{\rm C} \equiv \Theta^{(0)}_{\rm eq} \equiv \eta_{\Delta n^{(0)}_0} = \frac{\int x^3 \Delta n^{(0)}_0 \, w_y \, {\rm d}x}{4 E_{n_{\rm bb}}} \approx \Theta^{(0)}_0 + \sum_k \eta_{Y_k}\, y^{(0)}_{k,0} + \eta_M\, \mu^{(0)}_0, \tag{C.1b}$$

where $\rho_z = \frac{8\pi h}{c^3}\left(\frac{k_{\rm B} T_z}{h}\right)^4 E_{n_{\rm bb}}$ is the energy density of a blackbody at a temperature T_z and $\Theta_{\rm e} = \Delta T_{\rm e}/T_z$ denotes the relative temperature difference of $T_{\rm e}$ and T_z. For convenience we also introduced the background Comptonisation rate, κ, which will be used frequently below. We furthermore used the Compton integral, η_f, and the energy integral, $E_f = \int x^3 f(x)\, {\rm d}x$ as defined in paper I.

With Eq. (A.4), but setting $\Psi^{(1)} = 0$ to evaluate in the local inertia frame, we can also directly write the first order Compton energy exchange rate as

$$\begin{aligned}
\Lambda^{(1)}_C &\approx \kappa \left\{-\Theta^{(1)}_{\rm e} - \frac{\int x^3 \left[\hat{\mathcal{K}}_x \Delta n^{(1)}_0 + \Theta^{(1)}_{\rm e} \hat{\mathcal{D}}_x \Delta n^{(0)}_0\right] {\rm d}x}{4 E_{n_{\rm bb}}} - 4\Theta^{(1)}_0 \Theta^{(0)}_{\rm e} + \delta^{(1)}_{\rm b} \left[\Theta^{(0)}_{\rm C} - \Theta^{(0)}_{\rm e}\right]\right\} \\
&= \kappa \left\{\left[1 + \delta^{(0)}_{\gamma,0}\right]\left[\Theta^{(1)}_{\rm C} - \Theta^{(1)}_{\rm e}\right] + \left[\delta^{(1)}_{\rm b} + 4\Theta^{(1)}_0\right]\left[\Theta^{(0)}_{\rm C} - \Theta^{(0)}_{\rm e}\right]\right\}
\end{aligned} \tag{C.2a}$$

$$\Theta^{(1)}_{\rm C} = \frac{1}{1 + \delta^{(0)}_{\gamma,0}} \left[\frac{\int x^3 \Delta n^{(1)}_0 \, w_y \, {\rm d}x}{4 E_{n_{\rm bb}}} - 4\Theta^{(1)}_0\, \Theta^{(0)}_{\rm eq}\right] \approx \Theta^{(1)}_{\rm eq} - \Theta^{(1)}_0\, \delta^{(0)}_{\gamma,0} - 4\Theta^{(1)}_0\, \Theta^{(0)}_{\rm eq}. \tag{C.2b}$$

with $\delta^{(0)}_{\gamma,0} = E_{\Delta n^{(0)}_0}/E_{n_{\rm bb}} = 4\Theta^{(0)}_0 + 4\sum_n y^{(0)}_{0,n} + \epsilon_M \mu^{(0)}_0$, which accounts for a small correction to the energy density of the zeroth order photon field. We will see that this term ensures energy conservation with respect to the scattering correction, $\Delta\theta^{(1)}_{\rm e} \hat{\mathcal{D}}_x \Delta n^{(0)}_0$ in Eq. (A.4).

---

[8] The sign is chosen such that for a blackbody spectrum with $T_\gamma > T_{\rm e}$ one has $\Lambda^{(0)}_C > 0$ or energy flow to the electrons.

We comment that the Compton equilibrium temperature corrections, $\Theta_{\rm C}^{(0)}$ and $\Theta_{\rm C}^{(1)}$, could also have been obtained directly using the well-known expression [68, 69]

$$\Theta_{\rm C} = \frac{\int x^4 n(1+n)\,{\rm d}x}{4\int x^3 n\,{\rm d}x} - 1 = \frac{\int x^3 \Delta n_0\, w_y\,{\rm d}x + \int x^4 \Delta n_0^2\,{\rm d}x}{4E_{n_{\rm bb}} + 4\int x^3 \Delta n_0\,{\rm d}x} \tag{C.3a}$$

both at zeroth and first order in perturbation when non-linear distortion terms ($\propto \Delta n_0^2$) are neglected. The direct path using the Compton collision term highlights the consistency of the result.

C.2 Electron temperature equation and effective photon heating rate

In this section, we consider solutions to the local electron temperature including perturbations in the medium. Due to the presence of Compton scattering, the electron temperature is always pushed extremely close to the Compton equilibrium temperature, with corrections from the heating terms that appear in the electron temperature equation. This leads to an effective heating term in the photon equation that is obtained here for conditions in the pre-recombination era.

To obtain the correct terms, we adapt the discussion of [35]. Denoting time-derivatives in this context with 'dot', during the pre-recombination era and consistent up to first order in perturbations one has[9]

$$C_V\left[\dot{T}_{\rm e} + 2H\,T_{\rm e} + 2\left(\frac{\beta}{3a} + \dot{\Phi}\right)T_{\rm e}\right] = (1+\Psi)\left[\Lambda_C + \dot{Q}_{\rm c}\right] \tag{C.4}$$

for the electron temperature. Here, $C_V = \frac{3}{2}k_{\rm B}\,N_{\rm tot}$ is the non-relativistic heat capacity of the baryons, with $N_{\rm tot} = N_{\rm H} + N_{\rm He} + N_{\rm e}$ being the total number density of baryonic particles.

Armed with these ingredients we can now write down the evolution equation for the electron temperature at background and perturbed level. This yields

$$\begin{aligned} C_V^{(0)}\left[\dot{T}_{\rm e}^{(0)} + 2HT_{\rm e}^{(0)}\right] &= \Lambda_C^{(0)} + \dot{Q}_{\rm c}^{(0)},\\ C_V^{(1)}\left[\dot{T}_{\rm e}^{(0)} + 2HT_{\rm e}^{(0)}\right] + C_V^{(0)}\left[\dot{T}_{\rm e}^{(1)} + 2HT_{\rm e}^{(1)} + 2\left(\frac{\beta^{(1)}}{3a} + \dot{\Phi}^{(1)}\right)T_{\rm e}^{(0)}\right] &= \Psi^{(1)}\left[\Lambda_C^{(0)} + \dot{Q}_{\rm c}^{(0)}\right] + \Lambda_C^{(1)} + \dot{Q}_{\rm c}^{(1)}. \end{aligned}$$

Assuming the plasma to be fully ionised one finds $C_V^{(1)} = C_V^{(0)}\,\delta_{\rm b}^{(1)}$.

The background level equation is well-known in connection with the standard recombination and thermalisation problems [e.g., 18, 70]. In the pre-recombination era, $T_{\rm e}$ will follow a sequence of quasi-stationary states due to rapid Compton interactions, pushing $T_{\rm e}^{(0)} \approx T_z$. Making the Ansatz $T_{\rm e}^{(0)} = T_z + \Delta T_{\rm e}^{(0)}$ and setting $\Delta\dot{T}_{\rm e}^{(0)} \approx 0$, we find

$$\begin{aligned} C_V^{(0)}\left[-HT_z + 2HT_z\left(1 + \Theta_{\rm e}^{(0)}\right)\right] &\approx \kappa\left[\Theta_{\rm eq}^{(0)} - \Theta_{\rm e}^{(0)}\right] + \dot{Q}_{\rm c}^{(0)}.\\ \rightarrow \quad \Theta_{\rm e}^{(0)} = \frac{\Delta T_{\rm e}^{(0)}}{T_z} &\approx \frac{\kappa\Theta_{\rm eq}^{(0)} + \dot{Q}_{\rm c}^{(0)} - HC_V^{(0)}T_z}{\kappa + 2H\,C_V^{(0)}} \approx \Theta_{\rm eq}^{(0)} + \frac{\dot{Q}_{\rm c}^{(0)}}{\kappa}, \end{aligned} \tag{C.5}$$

where in the second step we neglected the correction from the adiabatic cooling effect, which leads to a tiny average distortion $\mu_{\rm cool} \simeq -3\times10^{-9}$ [28, 71, 72]. We could have directly obtained the solution by setting $\Lambda_C^{(0)} + \dot{Q}_{\rm c}^{(0)} \equiv \kappa\left[\Theta_{\rm eq}^{(0)} - \Theta_{\rm e}^{(0)}\right] + \dot{Q}_{\rm c}^{(0)} \approx 0$, a fact that we will use when computing the perturbed temperature solution. Starting with Eq. (C.5) for the electron temperature, we then obtain

$$\left.\frac{\partial \Delta n_0^{(0)}}{\partial t}\right|_{\rm K} = \dot{\tau}\theta_z\left[\Theta_{\rm e}^{(0)}Y + \hat{\mathcal{K}}_x\Delta n_0^{(0)}\right] = \dot{\tau}\theta_z\left[\Theta_{\rm eq}^{(0)}Y + \hat{\mathcal{K}}_x\Delta n_0^{(0)}\right] + \frac{\dot{Q}_{\rm c}^{(0)}}{4\rho_z}\,Y \tag{C.6}$$

---

[9] We use $\Psi = \Psi^{\rm S}$ and $\Phi = -\Phi^{\rm S}$.

for the Kompaneets terms. The combination $\Theta^{(0)}_{\rm eq} Y + \hat{\mathcal{K}}_x \Delta n^{(0)}_0$ can be replaced with the Kompaneets matrix description $\rightarrow M_K\, \boldsymbol{y}^{(0)}$. Here, $\Theta^{(0)}_{\rm eq} \approx \Theta^{(0)}_0 + \sum_k \eta_{Y_k}\, y^{(0)}_{k,0} + \eta_M\, \mu^{(0)}_0$ once Δn_0^2 terms are neglected. From this we can also identify the relative effective heating rate

$$\dot{Q}^{(0)} \equiv \frac{\dot{Q}^{(0)}_{\rm c}}{\rho_z}, \tag{C.7}$$

which then leads to the formulation for the background distortion evolution given in the main text.

We note that when exact Compton equilibrium is no longer reached, we can directly solve the zeroth order temperature equation given above to compute the main distortion source term in Eq. (C.6). In this regime, we can neglect the terms $\dot{\tau}\theta_z$ and simply find direct sources of y-distortions without any spectral evolution. This only is expected to become important at $z \lesssim 200$, which is a regime we do not consider at this point, such that the quasi-stationary solution is valid.

To obtain the perturbed temperature solution, we take the anticipated short-cut assuming quasi-stationary conditions (or more accurately, $C_V H/\kappa \ll 1$). This then yields

$$0 \approx \Psi^{(1)}\left[\Lambda^{(0)}_C + \dot{Q}^{(0)}_{\rm c}\right] + \Lambda^{(1)}_C + \dot{Q}^{(1)}_{\rm c}. \tag{C.8}$$

This expression neglects the perturbed corrections to the adiabatic cooling process, which are sub-dominant for our purposes, but should lead to a minimal distortion anisotropy in ΛCDM. In our limit, $\Lambda^{(0)}_C \approx \kappa\left[\Theta^{(0)}_{\rm eq} - \Theta^{(0)}_{\rm e}\right] \approx -\dot{Q}^{(0)}_{\rm c}$ and hence $\Lambda^{(0)}_C + \dot{Q}^{(0)}_{\rm c} \approx 0$. With Eq. (C.2), this then implies

$$\begin{aligned}
0 &\approx \kappa\left\{\left[1+\delta^{(0)}_{\gamma,0}\right]\left[\Theta^{(1)}_{\rm C} - \Theta^{(1)}_{\rm e}\right] + \left[\delta^{(1)}_{\rm b} + 4\Theta^{(1)}_0\right]\left[\Theta^{(0)}_{\rm eq} - \Theta^{(0)}_{\rm e}\right]\right\} + \dot{Q}^{(1)}_{\rm c} \\
&\approx \kappa\left[1+\delta^{(0)}_{\gamma,0}\right]\left[\Theta^{(1)}_{\rm C} - \Theta^{(1)}_{\rm e}\right] - \left[\delta^{(1)}_{\rm b} + 4\Theta^{(1)}_0\right]\dot{Q}^{(0)}_{\rm c} + \dot{Q}^{(1)}_{\rm c}
\end{aligned} \tag{C.9}$$

By solving for $\Theta^{(1)}_{\rm e}$ and neglecting higher order distortion terms, we then finally find

$$\Theta^{(1)}_{\rm e} \approx \Theta^{(1)}_{\rm C} + \frac{\dot{Q}^{(1)}_{\rm c}}{\kappa} - \left[\delta^{(1)}_{\rm b} + 4\Theta^{(1)}_0\right]\frac{\dot{Q}^{(0)}_{\rm c}}{\kappa} \approx \frac{\Theta^{(1)}_{\rm eq}}{1+\delta^{(0)}_{\gamma,0}} + \frac{\dot{Q}^{(1)}_{\rm c}}{\kappa} - 4\Theta^{(1)}_0\Theta^{(0)}_{\rm eq} - \left[\delta^{(1)}_{\rm b} + 4\Theta^{(1)}_0\right]\frac{\dot{Q}^{(0)}_{\rm c}}{\kappa}, \tag{C.10}$$

where the term $\propto 4\Theta^{(1)}_0$ accounts for variations of the local photon heat capacity. Putting everything together, from Eq. (A.4) we find

$$\begin{aligned}
\left.\frac{\partial \Delta n^{(1)}_0}{\partial t}\right|_{\rm K} &\approx \dot{\tau}\theta_z\left[\Theta^{(1)}_{\rm e} Y + \hat{\mathcal{K}}_x \Delta n^{(1)}_0 + \Theta^{(1)}_{\rm e}\hat{\mathcal{D}}_x \Delta n^{(0)}_0 + 4\Theta^{(1)}_0\Theta^{(0)}_{\rm e}\, Y_1\right] + \dot{\tau}\theta_z\left[\delta^{(1)}_{\rm b} + \Psi^{(1)}\right]\left[\Theta^{(0)}_{\rm e} Y + \hat{\mathcal{K}}_x \Delta n^{(0)}_0\right] \\
&\approx \dot{\tau}\theta_z\left[(\Theta^{(1)}_{\rm eq} - \Theta^{(1)}_0\delta^{(0)}_{\gamma,0})Y + \hat{\mathcal{K}}_x \Delta n^{(1)}_0 + 4\Theta^{(1)}_0\Theta^{(0)}_{\rm eq}\,(Y_1 - Y)\right] + \dot{\tau}\theta_z\left[\delta^{(1)}_{\rm b} + \Psi^{(1)}\right]\left[\Theta^{(0)}_{\rm eq} Y + \hat{\mathcal{K}}_x \Delta n^{(0)}_0\right] \\
&\qquad + \dot{\tau}\theta_z\Theta^{(1)}_0\hat{\mathcal{D}}_x \Delta n^{(0)}_0 + \left[\frac{\dot{Q}^{(1)}_{\rm c}}{\rho_z} + \Psi^{(1)}\,\frac{\dot{Q}^{(0)}_{\rm c}}{\rho_z}\right]\frac{Y}{4} + \Theta^{(1)}_0\,\frac{\dot{Q}^{(0)}_{\rm c}}{\rho_z}\,(Y_1 - Y) \\
&\approx \dot{\tau}\theta_z\left[\Theta^{(1)}_{\rm eq}\, Y + \hat{\mathcal{K}}_x \Delta n^{(1)}_0\right] + \dot{\tau}\theta_z\left[\delta^{(1)}_{\rm b} + \Psi^{(1)}\right]\left[\Theta^{(0)}_{\rm eq} Y + \hat{\mathcal{K}}_x \Delta n^{(0)}_0\right] + \left[\frac{\dot{Q}^{(1)}_{\rm c}}{\rho_z} + \Psi^{(1)}\,\frac{\dot{Q}^{(0)}_{\rm c}}{\rho_z}\right]\frac{Y}{4} \\
&\qquad + \dot{\tau}\theta_z\Theta^{(1)}_0\left(\hat{\mathcal{D}}_x \Delta n^{(0)}_0 - \delta^{(0)}_{\gamma,0}\, Y\right) + 4\dot{\tau}\theta_z\,\Theta^{(1)}_0\left[\Theta^{(0)}_{\rm eq} + \frac{\dot{Q}^{(0)}_{\rm c}}{\kappa}\right](Y_1 - Y)
\end{aligned} \tag{C.11}$$

for the Kompaneets terms at first order. We can identify $\Theta^{(1)}_{\rm eq} Y + \hat{\mathcal{K}}_x \Delta n^{(1)}_0 = M_K\, \boldsymbol{y}^{(1)}$ like for the zeroth order. We note that in the second line we used the equilibrium temperature differences to isolate the effect of external heating terms. We also set $\Theta^{(1)}_{\rm e}\hat{\mathcal{D}}_x \Delta n^{(0)}_0 \approx \Theta^{(1)}_0\hat{\mathcal{D}}_x \Delta n^{(0)}_0$, given that the other corrections to $\Theta^{(1)}_{\rm e}$ are higher order in the distortion.

It is important to highlight that the terms in the last line of Eq. (C.11) do not add any extra energy to the system, but merely redistribute the energy spectrally, leading to a correction to the perturbed scattering effect. For the terms proportional to $\propto Y_1 - Y$ this is trivially seen when computing $\int x^3 Y \,\mathrm{d}x = \int x^3 Y_1 \,\mathrm{d}x = 4E_{n_{\rm bb}}$. For the other term we have $\int x^3 \hat{\mathcal{D}}_x \Delta n_0^{(0)} \,\mathrm{d}x = -\int x^4 \partial_x \Delta n_0^{(0)} \,\mathrm{d}x = 4\int x^3 \Delta n_0^{(0)} \,\mathrm{d}x = 4E_{n_{\rm bb}}\,\delta^{(0)}_{\gamma,0}$. With $\int x^3 Y \,\mathrm{d}x = 4E_{n_{\rm bb}}$ this identically chancels the remaining term in the last line and also highlights how the term $\propto \delta^{(0)}_{\gamma,0}$ is crucial for conserving energy.

The only terms in Eq. (C.11) leading to addition of energy are those $\propto \dot{Q}_{\rm c}^{(1)}/\rho_z + \Psi^{(1)}\,\dot{Q}_{\rm c}^{(0)}/\rho_z$. We thus identify them as

$$\dot{Q}^{(1)} \equiv \frac{\dot{Q}_{\rm c}^{(1)}}{\rho_z} + \Psi^{(1)}\frac{\dot{Q}_{\rm c}^{(0)}}{\rho_z}, \tag{C.12}$$

which is used in the main text to formulate the first order heating rate. We highlight that the terms $\dot{Q}_{\rm c}^{(0)}$ and $\dot{Q}_{\rm c}^{(1)}$ simply follow from the collision terms in the local inertial frame. One would have naturally guessed this perturbed heating term when thinking about $\dot{Q} \approx (1+\Psi)\dot{Q}_{\rm c}/\rho_z$.

C.2.1 Adding extra Doppler terms to the system

Since the terms in the last line of Eq. (C.11) do not add any extra energy (or photons) to the system, we can in principle neglect them in our computation, without changing the consistency of the system. However, it is fairly easy to include them once $\hat{\mathcal{D}}_x \Delta n_0^{(0)}$ is evaluated. Since they are expected to change the exact distortion evolution across the residual era, this could provide extra time-dependent information. Inserting our distortion representation, with $\hat{\mathcal{D}}_x = -3\hat{\mathcal{O}}_x + \hat{\mathcal{O}}_x^2$ we have

$$\begin{aligned}
&\hat{\mathcal{D}}_x G = 4Y_1, \qquad \hat{\mathcal{D}}_x Y_n = 12Y_{n+1} + 16\,Y_{n+2}, \qquad \hat{\mathcal{D}}_x \Delta n_0^{(0)} = 4\Theta_0^{(0)}\,Y_1 + \hat{\mathcal{D}}_x \Delta n_{0,d}^{(0)} \\
&\hat{\mathcal{D}}_x \Delta n_{0,\rm d}^{(0)} = 4\sum_{n=0}^{N-2}(3Y_{n+1} + 4Y_{n+2})\,y_{0,n}^{(0)} + \hat{\mathcal{D}}_x Y_{N-1} y_{0,N-1}^{(0)} + \hat{\mathcal{D}}_x Y_N\, y_{0,N}^{(0)} + \hat{\mathcal{D}}_x M\,\mu_0^{(0)},
\end{aligned} \tag{C.13}$$

where we separated the distortion only terms (i.e., those without G). We can in principle again use our projection method to obtain descriptions of $\hat{\mathcal{D}}_x Y_{N-1}$, $\hat{\mathcal{D}}_x Y_N$ and $\hat{\mathcal{D}}_x M$ in our computation basis, but it is much easier to just apply the boost matrix, $M_{\rm B}$, defined in Eq. (2.9) for this purpose. It turns out that we then have

$$\hat{\mathcal{D}}_x \Delta n_0^{(0)} = (M_{\rm B} - 3I)M_{\rm B}\,\boldsymbol{y}^{(0)} = M_{\rm D}\,\boldsymbol{y}^{(0)} \tag{C.14}$$

where I is the identity matrix. One can precompute the *Doppler matrix* $M_{\rm D} = (M_{\rm B} - 3I)M_{\rm B}$ for numerical applications to ease the computations.

C.2.2 Compton equilibrium spectrum

An important property of the Compton collision term operator is the associated equilibrium spectrum, which leads to a stationary spectrum under repeated scatterings. At zeroth order in perturbations, we trivially have $\Delta n_{0,\rm eq}^{(0)} = G\,\Theta_0^{(0)} + M\,\mu_0^{(0)}$, as can be verified by using $\hat{\mathcal{K}}_x n_{0,\rm eq}^{(0)} = -Y\,(\Theta_0^{(0)} + \eta_M\,\mu_0^{(0)})$ and $\Theta_{\rm eq}^{(0)} = \Theta_0^{(0)} + \eta_M\,\mu_0^{(0)}$, which leads to cancellation of the Kompaneets term at zeroth order.

How does this work for the first order Comptonisation terms? If we assume no average distortion or temperature shift and no heating, then from Eq. (C.11) we naturally have

$$\left.\frac{\partial \Delta n_0^{(1)}}{\partial t}\right|_{\rm K} \approx \dot{\tau}\theta_z\left[\Theta_{\rm eq}^{(1)}\,Y + \hat{\mathcal{K}}_x \Delta n_0^{(1)}\right], \tag{C.15}$$

which again has the general solution $\Delta n^{(1)}_{0,\rm eq} = G\,\Theta^{(1)}_0 + M\,\mu^{(1)}_0$ and $\Theta^{(1)}_{\rm eq} = \Theta^{(1)}_0 + \eta_M\,\mu^{(1)}_0$. Due to the absence of any source terms for M we would even expect $\mu^{(1)}_0 = 0$. Comptonisation would therefore not modify the spectrum of temperature perturbations, as naturally expected.

Let us now assume that the average spectrum has a slightly higher temperature than T_z, i.e., $\Delta n^{(0)}_{0,\rm eq} = G\,\Theta^{(0)}_0$. In the absence of external heating, from Eq. (C.11) we find

$$\left.\frac{\partial \Delta n^{(1)}_0}{\partial t}\right|_{\rm K} \approx \dot{\tau}\theta_z\left[\Theta^{(1)}_{\rm eq}\,Y + \hat{\mathcal{K}}_x \Delta n^{(1)}_0\right] + 8\dot{\tau}\theta_z\,\Theta^{(1)}_0\Theta^{(0)}_0\,(Y_1 - Y) \tag{C.16}$$

Where we used $\hat{\mathcal{D}}_x \Delta n^{(0)}_0 - \delta^{(0)}_{\gamma,0}\,Y = 4\Theta^{(0)}_0(Y_1 - Y)$. The first group of terms is again solved as before, but the last term remains out of equilibrium. It neither adds energy nor photon number to the spectrum, but just changes the spectral shape. This means that even if the average spectrum has reached a thermal spectrum, the anisotropy spectrum is distorted unless photon production and redistribution are very efficient. We remark that this conclusion does not change when adding back the omitted stimulated scattering term, $2\theta_z \hat{\mathcal{D}}^*_x \Delta n^{(0)}_0 \Delta n^{(1)}_0$.

C.3 Photon production in the μ-era

A pure μ-distortion spectrum, M, does not balance emission and absorption at low frequencies, leading to net photon production in the μ-era and slow evolution of the μ-parameter at high frequencies. However, if we insert $\Delta n_0 = G(x)\,\Theta_0 + M(x)\,\mu_0$ into the collision term and compute the photon production rate, we obtain a divergent result at low frequencies. We thus need to find a modified solution for the distortion that fixes this problem.

C.3.1 Zeroth order treatment of photon production

If we consider the photon evolution equation at zeroth order, with Eq. (C.6) and (A.10a) we have

$$\frac{\partial \Delta n^{(0)}_0}{\partial \tau} \approx \theta_z\left[\Theta^{(0)}_{\rm eq} Y + \hat{\mathcal{K}}_x \Delta n^{(0)}_0\right] - \frac{\Lambda(x,\theta_z)}{x^2}\,\frac{n_{\rm bb}}{G}\left[\Delta n^{(0)}_0 - \Theta^{(0)}_{\rm eq}\,G\right], \tag{C.17}$$

where we used $(1-{\rm e}^{-x})/x = {\rm e}^{-x}({\rm e}^{x}-1)/x = n_{\rm bb}/G$ and assumed that the heating process led to a non-zero initial chemical potential and temperature shift at the initial time. Clearly, a spectrum $\Delta n^{(0)}_0 = \Theta^{(0)}_{\rm eq}\,G$ is a stationary solution of the problem. The strategy is now to assume quasi-stationary evolution of the spectrum, which means the left hand side is set to zero, $\partial_\tau \Delta n^{(0)}_0 \approx 0$. Without emission and absorption terms we know that the spectrum will not change in this case and simply be given by $\Delta n^{(0)}_0 = G(x)\,\Theta^{(0)}_0 + M(x)\,\mu^{(0)}_0$ with $\Theta^{(0)}_0$ and $\mu^{(0)}_0$ fixed by the initial condition. However, as already mentioned, at low frequency this does not solve the equation above. Energetically the required correction does not matter much, which means we can still use $\Theta^{(0)}_{\rm eq} \approx \Theta^{(0)}_0 + \eta_M\,\mu^{(0)}_0$.

To obtain the correction to the spectrum, we take the low-frequency limit of Eq. (C.17), essentially thinking of the problem as a high-frequency (↔ energy) evolution and low-frequency (↔ photon number) evolution [e.g., see 19, for more details]. The characteristic scale in the problem is the *critical frequency*, $x_{\rm c} \ll 1$, which is defined by the balance between photon emission processes and Compton scattering [15, 16, 18, 19, 71, 73]. This frequency is determined (at leading order) by the implicit equation $\Lambda(x_{\rm c},\theta_z) \approx \theta_z x_{\rm c}^2$, which assumes that the emission coefficient, $\Lambda(x_{\rm c},\theta_z)$, is a slowly varying function of x. In the DC-era, we have $x_{\rm c} \approx 8.6\times10^{-4}\,\sqrt{(1+z)/2\times10^6}$, assuming the standard CMB temperature for T_z. The idea is now to write the Ansatz $\Delta n^{(0)}_0 = G(x)\,\Theta^{(0)}_0 g(x) + M(x)\,\mu^{(0)}_0 m(x)$ with frequency-dependent correction functions $g(x)$ and $m(x)$. At $x \gg x_{\rm c}$ one has $g(x) \simeq m(x) \simeq 1$. It is therefore useful to consider $g(x)$ and $m(x)$ as functions of $\xi = x_{\rm c}/x$ instead of x. Starting from Eq. (C.17), multiplying it by $-x_\mathrm{c}^2/[\theta_z \xi^4]$ (this is the leading order dependence) and replacing $x \to x_\mathrm{c}/\xi$, with the Ansatz $\Delta n_0^{(0)} = G(x_\mathrm{c}/\xi)\,\Theta_0^{(0)} g(\xi) + M(x_\mathrm{c}/\xi)\,\mu_0^{(0)} m(\xi)$ we then find

$$\begin{aligned} 0 \approx \mu_0^{(0)} \left[\partial_\xi^2 m(\xi) - m(\xi)\right] &+ x_\mathrm{c}\mu_0^{(0)} \left\{\frac{1}{2}\frac{m(\xi)}{\xi} + \eta_M \left[\frac{2-\xi^2}{\xi^3} - \left(\frac{\partial^2}{\partial\xi^2}\frac{m(\xi)}{\xi} - \frac{m(\xi)}{\xi}\right)\right]\right\} \\ &+ x_\mathrm{c}\Theta_0^{(0)} \left\{\frac{2-\xi^2}{\xi^3}[1-g(\xi)] + \frac{4}{\xi^2}\,\partial_\xi g(\xi)\right\}. \end{aligned} \tag{C.18}$$

up to first order in $x_\mathrm{c} \ll 1$. Neglecting the terms $\propto x_\mathrm{c}$, we have the physical solutions $m(\xi) = \exp(-\xi)$ and $g(\xi) = 1$, which together gives

$$G^\mathrm{mod}(x) \approx G(x), \qquad M^\mathrm{mod}(x) \approx M(x)\,\mathrm{e}^{-x_\mathrm{c}/x} \tag{C.19}$$

for the modified temperature and μ-distortion spectrum. The goal is now use this to obtain an expression for the photon production rate at lowest order in x_c. Integrating the collision term in terms of photon number[10], we find

$$\begin{aligned} \frac{1}{N_z}\frac{\partial N_\mathrm{c}^{(0)}}{\partial\tau} &= -\frac{\int x^2\,\mathrm{d}x\,\frac{\Lambda(x,\theta_z)}{x^2}\,\frac{n_\mathrm{bb}}{G}\left[\Delta n_0^{(0)} - \Theta_\mathrm{eq}^{(0)}\,G\right]}{\int x^2 n_\mathrm{bb}(x)\,\mathrm{d}x} \approx -\frac{\theta_z}{N_{n_\mathrm{bb}}}\int_0^\infty x_\mathrm{c}^2\left[\Delta n_0^{(0)} - \Theta_\mathrm{eq}^{(0)}\,G\right]\mathrm{d}x \\ &= -\frac{\theta_z}{N_{n_\mathrm{bb}}}\int_0^\infty x_\mathrm{c}^2\left[G\,\Theta_0^{(0)} + M\,m\,\mu_0^{(0)} - G(\Theta_0^{(0)} + \eta_M\,\mu_0^{(0)})\right]\mathrm{d}x \\ &\approx \frac{\theta_z}{N_{n_\mathrm{bb}}}\int_0^\infty \frac{x_\mathrm{c}^2}{x^2}\,\mu_0^{(0)}\,m(x)\,\mathrm{d}x \\ &= \frac{\theta_z\,\mu_0^{(0)}}{N_{n_\mathrm{bb}}}\int_0^\infty \frac{x_\mathrm{c}^2}{x^2}\,\mathrm{e}^{-x_\mathrm{c}/x}\,\mathrm{d}x = \left(\frac{3}{4\,\gamma_\rho}\right)\gamma_N\,\theta_z\,x_\mathrm{c}\,\mu_0^{(0)}. \end{aligned} \tag{C.20}$$

from the collision term. This remarkable result is the key to understanding the rate of conversion from $\mu_0^{(0)} \to \Theta_0^{(0)}$. Indeed, the reduction of the chemical potential amplitude is given by $\partial_t\mu_0^{(0)}\big|_\mathrm{em/abs} \approx -\gamma_N\,\dot{\tau}\theta_z\,x_\mathrm{c}\,\mu_0^{(0)}$ [19], which we use in main text. We highlight again that without the low-frequency modification to the spectrum by the factor $\mathrm{e}^{-x_\mathrm{c}/x}$ one would not have obtained a finite photon production rate. At low frequencies, the distortion vanishes and the spectrum is in full equilibrium with the electrons, stopping any emission and absorption.

We note that, while for the temperature term $\propto \Theta_0^{(0)}$ an exact cancellation occurred, in the final steps we had to drop the term $\eta_M\mu_0^{(0)}$, since it still diverges logarithmically towards $x \to 0$. This is because we did not include higher orders in x_c. We can in principle use Eq. (C.18) to obtain correction to the distortion spectra at first order in x_c. However, the procedure becomes quite involved requiring additional normalisation conditions and also the including of time-dependent corrections [see 19]. This is beyond the scope of this work and left to a future publication.

We also note that one can in principle evaluate the exact integral $\int x_\mathrm{c}^2 M(x) m(x)\,\mathrm{d}x$ numerically. This yields the correction $x_\mathrm{c} \to x_\mathrm{c}(1 - 1.65x_\mathrm{c}^{0.88})$, which indicates that photon production is a little slower than in the soft photon limit. However, in terms of perturbations in x_c this is not a consistent treatment and other corrections are expected to have a similar level. We therefore do not recommend adding these corrections until a more complete treatment of the problem is attempted.

---

[10] There is no contribution from the scattering terms.

C.3.2 First order treatment of photon production

We can now repeat the same computation but at first order in perturbations. The emission terms at first order in perturbations is given in Eq. (A.10b). Remembering that $\Delta n_0^{(0)} \approx -\mu_0^{(0)}\, m(x)/x^2$ gave the leading order term at zeroth order and checking the coefficients of all functions, keeping only leading order terms, we then find

$$\left.\frac{\partial n_0^{(1)}}{\partial \tau}\right|_{\rm em/abs}^{\rm low} \approx \frac{\theta_z x_{\rm c}^2}{x^2}\left[\mu_0^{(1)}\,\frac{m^{(1)}(x)}{x^2} + \left(\delta_{\rm b}^{(1)} + \Psi^{(1)} + \Theta_0^{(1)}\left[\left.\frac{\partial \ln \Lambda}{\partial \ln \theta_\gamma}\right|_{\theta_z} + \left.\frac{\partial \ln \Lambda}{\partial \ln \theta_{\rm e}}\right|_{\theta_z} - 1\right]\right)\mu_0^{(0)}\,\frac{m^{(0)}(x)}{x^2}\right].$$

Assuming $\Lambda \propto \theta_\gamma^2$, we have $\frac{\partial \ln \Lambda}{\partial \ln \theta_\gamma} = 2$. Realising that $Y_1 - Y \approx 3/[2x]$ and $Y \approx -2/x$ at $x \ll 1$, we can also collect all leading order Compton terms in Eq. (C.11) as

$$\left.\frac{\partial \Delta n_0^{(1)}}{\partial \tau}\right|_{\rm K}^{\rm low} \approx \theta_z\left\{\hat{\mathcal{K}}_x^{\rm low}\Delta n_0^{(1)} + \left[\delta_{\rm b}^{(1)} + \Psi^{(1)}\right]\hat{\mathcal{K}}_x^{\rm low}\Delta n_0^{(0)} + \Theta_0^{(1)}\hat{\mathcal{D}}_x\Delta n_0^{(0)} + 2\hat{\mathcal{D}}_x^*\Delta n_0^{(0)}\Delta n_0^{(1)}\right\}$$

with $\hat{\mathcal{K}}_x^{\rm low} = x^{-2}\partial_x x^4[\partial_x{+}2/x] \equiv x^{-2}\partial_x x^2\partial_x x^2$. We added the term $2\hat{\mathcal{D}}_x^*\Delta n_0^{(0)}\Delta n_0^{(1)} \approx 2\Theta_0^{(1)}\hat{\mathcal{D}}_x^*\Delta n_0^{(0)}G(x)$, which naturally cancels some of the logarithmic corrections at low frequencies. However, in the overall evolution at high frequencies it should not be as important.

From the first order computation it is clear that the term $\propto \delta_{\rm b}^{(1)} + \Psi^{(1)}$ will be canceled by the corresponding zeroth order emission term. Inserting $\Delta n_0^{(i)} \approx -\mu_0^{(i)}\, m^{(i)}(x)/x^2$ and adding the emission terms we then obtain

$$0 \approx \mu_0^{(1)}\,\hat{\mathcal{K}}_x^{\rm low}\frac{m^{(1)}(x)}{x^2} + \Theta_0^{(1)}\,\mu_0^{(0)}\left[\hat{\mathcal{D}}_x\frac{m^{(0)}(x)}{x^2} + \hat{\mathcal{D}}_x^*\frac{m^{(0)}(x)}{x^3}\right] - \frac{x_{\rm c}^2}{x^4}\left[\mu_0^{(1)}\,m^{(1)}(x) + \Theta_0^{(1)}\,\mu_0^{(0)}\,m^{(0)}(x)\right].$$

after dividing through by $-\theta_z$. Transforming to $\xi = x_{\rm c}/x$ with $\partial_x = -(x_{\rm c}/x^2)\partial_\xi$ we then have $\hat{\mathcal{K}}_x^{\rm low} = (x_{\rm c}^2/x^4)\,\partial_\xi^2 x^2$, $\hat{\mathcal{D}}_x x^{-2} = (x_{\rm c}^2/x^4)\,\partial_\xi\xi^{-2}\partial_\xi\xi^2$ and $\hat{\mathcal{D}}_x^* x^{-3} = -(x_{\rm c}^2/x^4)\,\partial_\xi\xi^{-1}$ such that $\hat{\mathcal{D}}_x m^{(0)}(x)/x^2 = (x_{\rm c}^2/x^4)[1 - 2/\xi - 2/\xi^2]m^{(0)}(\xi)$ and $\hat{\mathcal{D}}_x^* m^{(0)}(x)/x^3 = (x_{\rm c}^2/x^4)[2/\xi + 2/\xi^2]m^{(0)}(\xi)$

$$\begin{aligned} 0 &\approx \mu_0^{(1)}\left[\partial_\xi^2 m^{(1)}(\xi) - m^{(1)}(\xi)\right] + \Theta_0^{(1)}\,\mu_0^{(0)}\left[\partial_\xi\xi^{-2}\partial_\xi\xi^2 m^{(0)}(\xi) - \partial_\xi\xi^{-1}\,m^{(0)}(\xi) - m^{(0)}(\xi)\right] \\ &\approx \mu_0^{(1)}\,m^{(0)}(\xi)\left[\partial_\xi^2 f - 2\partial_\xi f\right]. \end{aligned} \tag{C.21}$$

where in the last line we made the Ansatz $m^{(1)}(\xi) = m^{(0)}(\xi)\,f(\xi)$. We then obtain $f = 1$ as the main physical solution. This means that the first order perturbed distortion spectrum is identical to that of the zeroth order and highlights that the Comptonisation and emission correction terms cancel each other. We can therefore directly write the full photon production term as

$$\frac{1}{N_z}\frac{\partial N_{\rm c}^{(1)}}{\partial \tau} \approx \left(\frac{3}{4\gamma_\rho}\right)\gamma_N\,\theta_z\,x_{\rm c}\left[\mu_0^{(1)} + \left(\delta_{\rm b}^{(1)} + \Psi^{(1)} + \Theta_0^{(1)}\right)\mu_0^{(0)}\right]. \tag{C.22}$$

In addition to the expected term $-\gamma_N\,\dot{\tau}\theta_z\,x_{\rm c}\,\mu_0^{(1)}$, we therefore have the correction $-\gamma_N\,\dot{\tau}\theta_z\,x_{\rm c}\left(\delta_{\rm b}^{(1)} + \Psi^{(1)} + \Theta_0^{(1)}\right)\mu_0^{(0)}$, yielding a total reduction rate of

$$\partial_t\mu_0^{(1)} \approx -\gamma_N\,\dot{\tau}\theta_z\,x_{\rm c}\left[\mu_0^{(1)} + \left(\delta_{\rm b}^{(1)} + \Psi^{(1)} + \Theta_0^{(1)}\right)\mu_0^{(0)}\right]. \tag{C.23}$$

Following our approach for the zeroth order evolution, this can be used to model the redistribution of energy between $\mu_0^{(1)}$ and $\Theta_0^{(1)}$.

In the μ-era, we have a photon production rate $\simeq \theta_\gamma x_\mathrm{c} \propto \theta_\gamma^{3/2}$, where the local photon temperature matters. If the temperature changes, then both photon diffusion and DC/BR emission find a new equilibrium, naively yielding a perturbation to the photon production term $\simeq [\theta_\gamma x_\mathrm{c}(\theta_\gamma)]^{(1)} \propto \theta_z x_\mathrm{c}(\theta_z)\,[(1+\Theta_0)^{3/2}]^{(1)} \approx \theta_z x_\mathrm{c}(\theta_z)\,(3/2)\,\Theta_0^{(1)}$, as also used in Sect. 2.4.1. Why is the coefficient in front of $\Theta_0^{(1)}$ equal to 1? The reason is that in comparison to the zeroth order treatment we have three unexpected terms appearing in our quasi-stationary equation. For the emission process, this is the term $-\theta_z(x_\mathrm{c}^2/x^4)\mu_0^{(0)} m^{(0)}(x)$, while for the Compton contribution they are $\theta_z \Theta_0^{(1)} \hat{\mathcal{D}}_x \Delta n_0^{(0)} + 2\theta_z \hat{\mathcal{D}}_x^* \Delta n_0^{(0)} \Delta n_0^{(1)}$. Neglecting all these terms we have the quasi-stationary equation

$$0 \approx \mu_0^{(1)}\,\hat{\mathcal{K}}_x^{\rm low}\,\frac{m^{(1)}(x)}{x^2} - \frac{x_\mathrm{c}^2}{x^4}\left[\mu_0^{(1)}\,m^{(1)}(x) + 2\Theta_0^{(1)}\,\mu_0^{(0)}\,m^{(0)}(x)\right].$$

Carrying out the transformations as above, this then yields the equation

$$0 \approx \left\{\mu_0^{(1)}\left[\partial_\xi^2 f - 2\partial_\xi f\right] - 2\Theta_0^{(1)}\,\mu_0^{(0)}\right\} m^{(0)}(\xi),$$

which has the physical solution $f = 1 - \Theta_0^{(1)}\,\mu_0^{(0)}\,\xi/\mu_0^{(1)}$ or $\mu_0^{(1)} m_0^{(1)}(\xi) \approx (\mu_0^{(1)} - \Theta_0^{(1)}\,\mu_0^{(0)}\,\xi)\,m_0^{(0)}(\xi)$. Evaluating the photon production term with this correction, we find a contribution $-\Theta_0^{(1)}\mu_0^{(0)}$ from the terms $\mu_0^{(1)}\,m^{(1)}(\xi)$, which put together again yields the correct result, $-\Theta_0^{(1)}\,\mu_0^{(0)} + 2\Theta_0^{(1)}\,\mu_0^{(0)} = \Theta_0^{(1)}\,\mu_0^{(0)}$.

If we now only neglect the Compton terms $\theta_z \Theta_0^{(1)} \hat{\mathcal{D}}_x \Delta n_0^{(0)} + 2\theta_z \hat{\mathcal{D}}_x^* \Delta n_0^{(0)} \Delta n_0^{(1)}$, we find

$$0 \approx \left\{\mu_0^{(1)}\left[\partial_\xi^2 f - 2\partial_\xi f\right] - \Theta_0^{(1)}\,\mu_0^{(0)}\right\} m^{(0)}(\xi).$$

yielding $\mu_0^{(1)} m_0^{(1)}(\xi) \approx (\mu_0^{(1)} - \frac{1}{2}\Theta_0^{(1)}\,\mu_0^{(0)}\,\xi)\,m_0^{(0)}(\xi)$ and hence a contribution $-\Theta_0^{(1)}\,\mu_0^{(0)}/2$ from the production integral over $\mu_0^{(1)}\,m^{(1)}(\xi)$. Put together this then gives $-\Theta_0^{(1)}\,\mu_0^{(0)}/2 + \Theta_0^{(1)}\,\mu_0^{(0)} = (1/2)\,\Theta_0^{(1)}\,\mu_0^{(0)}$, which now is short of the expected result. Indeed, if we would only have used $[x_\mathrm{c}]^{(1)} \simeq [(1+\Theta_0)^{1/2}]^{(1)}$, this would have been the result, highlighting the link to the emission process only.[11]

If we only neglect the extra photon emission term, $-\theta_z(x_\mathrm{c}^2/x^4)\mu_0^{(0)} m^{(0)}(x)$, we *again* obtain

$$0 \approx \left\{\mu_0^{(1)}\left[\partial_\xi^2 f - 2\partial_\xi f\right] - \Theta_0^{(1)}\,\mu_0^{(0)}\right\} m^{(0)}(\xi).$$

and hence a contribution $-\Theta_0^{(1)}\,\mu_0^{(0)}/2$ from the production integral over $\mu_0^{(1)}\,m^{(1)}(\xi)$. But now adding things together we have $-\Theta_0^{(1)}\,\mu_0^{(0)}/2 + 2\Theta_0^{(1)}\,\mu_0^{(0)} = (3/2)\,\Theta_0^{(1)}\,\mu_0^{(0)}$. This suggests that the naive estimate does not capture the full effect of changing the DC emissivity with temperature.

While this is somewhat satisfying, we note that the required effect stems from a higher order correction caused by the absorption term $\propto -\Delta n_0\,(\mathrm{e}^{x\,\theta_z/\theta_\mathrm{e}} - 1)$ in Eq. (A.9), which does not contribute at zeroth order, and hence can also not be captured by perturbing the zeroth order equation. In reality, additional corrections can be expected since the frequency dependence of the DC emissivity, Λ, may also modify matters. This was highlighted before as part of the analytic computations for the distortion visibility function [19, 71]. Overall this shows that the precise scaling may depend on which approximation for the photon emissivity is indeed used. However, a more detailed discussion is beyond the scope of this work.

---

[11] The extra factor θ_γ in the photon production term $\propto \theta_\gamma x_\mathrm{c}$ comes from Compton scattering only.

D Details of the derivation for the line-of-sight integral solution.

To simplify Eq. (2.32) and obtain an expression for $\tilde{\boldsymbol{y}}_\ell^{(1)}(\eta_f, k)$ at the conformal time η_f, we follow the standard steps [57]. Defining the Thomson optical depth as $\tau = \int_0^\eta \tau'(\eta')\,\mathrm{d}\eta'$, it then follows

$$\partial_\eta \boldsymbol{y}^{(1)} + \mathbb{i}k\chi\,\boldsymbol{y}^{(1)} + \tau'\boldsymbol{y}^{(1)} = \mathrm{e}^{-(\mathbb{i}k\chi\eta+\tau)}\partial_\eta\left[\mathrm{e}^{\mathbb{i}k\chi\eta+\tau}\,\boldsymbol{y}^{(1)}\right]. \tag{D.1}$$

We can therefore formally integrate Eq. (2.32) [57]

$$\boldsymbol{y}^{(1)}(\eta_f, \chi, k) = \int_0^{\eta_f} \mathrm{d}\eta\,\mathrm{e}^{-\mathbb{i}k\chi\Delta\eta-\tau_\mathrm{b}}\,\boldsymbol{S}_\mathrm{LOS}(\eta, \chi, k) \tag{D.2}$$

to obtain the solution at the final time η_f. Here we defined $\tau_\mathrm{b} = \tau(\eta_f) - \tau(\eta)$ and $\Delta\eta = \eta_f - \eta$ for convenience. Note that now $\partial_\eta\tau_\mathrm{b} = -\tau'$.

In the source term, three types angular dependencies appear:

$$\begin{aligned}
\boldsymbol{S}_{\mathrm{LOS},\,0} &= \tau'\tilde{\boldsymbol{y}}_0^{(1)} - \boldsymbol{b}_0^{(0)}\frac{\partial\tilde{\Phi}^{(1)}}{\partial\eta} + \frac{\boldsymbol{Q}'^{(1)}}{4}, &\text{(D.3a)}\\
\boldsymbol{S}^{\mathrm{therm}}_{\mathrm{LOS},\,0} &= \tau'\theta_z\left[M_\mathrm{K}\,\tilde{\boldsymbol{y}}_0^{(1)} + \boldsymbol{D}_0^{(1)} + \left[\tilde{\delta}_\mathrm{b}^{(1)} + \tilde{\Psi}_\mathrm{b}^{(1)}\right]\left(M_\mathrm{K}\,\boldsymbol{y}_0^{(0)} + \boldsymbol{D}_0^{(0)}\right) + \tilde{\Theta}_0^{(1)}\,\boldsymbol{D}_0^{(0)}\right] &\text{(D.3b)}\\
\boldsymbol{S}_{\mathrm{LOS},\,1} &\equiv -\boldsymbol{b}_0^{(0)}\mathbb{i}k\chi\Psi^{(1)} + \tau'\beta^{(1)}\chi\,\boldsymbol{b}_0^{(0)} = -\mathbb{i}k\chi\left[\tilde{\Psi}^{(1)} + \frac{\tau'}{k}\tilde{\beta}^{(1)}\right]\boldsymbol{b}_0^{(0)}, &\text{(D.3c)}\\
\boldsymbol{S}_{\mathrm{LOS},\,2} &\equiv \frac{\tau'}{10}\,\boldsymbol{y}_2^{(1)} = -\frac{\tau'}{2}\,\tilde{\boldsymbol{y}}_2^{(1)}\,P_2(\chi). &\text{(D.3d)}
\end{aligned}$$

Here we separated thermalisation terms $\propto \tau'\theta_z$ from other sources. Inserting this back into Eq. (D.2) and carrying out the Legendre transform, we then have

$$\begin{aligned}
\tilde{\boldsymbol{y}}_\ell^{(1)}(\eta_f, k) &= \frac{\mathbb{i}^\ell}{2}\int P_\ell(\chi)\,\boldsymbol{y}^{(1)}(\eta_f, \chi, k)\,\mathrm{d}\chi = \int_0^{\eta_f}\mathrm{d}\eta\,g(\eta)\,\tilde{\boldsymbol{S}}_\ell(\eta, \eta_f, k), &\text{(D.4a)}\\
\tilde{\boldsymbol{S}}_\ell(\eta, \eta_f, k) &= \frac{\mathbb{i}^\ell}{2}\int P_\ell(\chi)\,\mathrm{e}^{-\mathbb{i}k\chi\Delta\eta}\,\frac{\boldsymbol{S}_\mathrm{LOS}(\eta, \chi, k)}{\tau'}\,\mathrm{d}\chi, &\text{(D.4b)}
\end{aligned}$$

where we introduced the visibility function, $g(\eta) = \tau'\,\mathrm{e}^{-\tau_\mathrm{b}} = \partial_\eta\mathrm{e}^{-\tau_\mathrm{b}}$. Using the identity [37]

$$\mathrm{e}^{-\mathbb{i}k\chi\Delta\eta} = \sum_\ell(-\mathbb{i})^\ell(2\ell+1)\,j_\ell[k\Delta\eta]\,P_\ell(\chi), \tag{D.5}$$

in terms of spherical bessel function, $j_\ell(x)$, we then encounter the following cases

$$\begin{aligned}
&\frac{\mathbb{i}^\ell}{2}\int P_\ell(\chi)\,\mathrm{e}^{-\mathbb{i}k\chi\Delta\eta}\,\mathrm{d}\chi = j_\ell(k\Delta\eta), &\text{(D.6a)}\\
&\frac{\mathbb{i}^\ell}{2}\int P_\ell(\chi)\,(-\mathbb{i}k\chi)\,\mathrm{e}^{-\mathbb{i}k\chi\Delta\eta}\,\mathrm{d}\chi = \partial_{\Delta\eta}j_\ell(k\Delta\eta) = k\left[\frac{\ell}{2\ell+1}j_{\ell-1}(k\Delta\eta) - \frac{\ell+1}{2\ell+1}j_{\ell+1}(k\Delta\eta)\right], &\text{(D.6b)}\\
&\frac{\mathbb{i}^\ell}{2}\int P_\ell(\chi)\,[-P_2(\chi)]\,\mathrm{e}^{-\mathbb{i}k\Delta\chi\eta}\,\mathrm{d}\chi = \frac{j_\ell(k\Delta\eta)}{2} + \frac{3}{2}\,\partial^2_{k\Delta\eta}j_\ell(k\Delta\eta), &\text{(D.6c)}
\end{aligned}$$

in Eq. (D.4b). Putting things together we then find $\tilde{\boldsymbol{S}}_\ell(\eta, \eta_f, k)$ as given in Eq. (2.33).[12]

---

[12] As an intermediate step we used that $\int g(\eta)\,\frac{k}{\tau'}\,\tilde{\Psi}^{(1)}\,j'_\ell(k\eta)\,\mathrm{d}\eta = \int g(\eta)\,\frac{1}{\tau'}\,[\partial_\eta\tilde{\Psi}^{(1)}]\,j_\ell(k\eta)\,\mathrm{d}\eta + \int g(\eta)\tilde{\Psi}^{(1)}\,j_\ell(k\eta)\,\mathrm{d}\eta$ since $k\tilde{\Psi}^{(1)}\,j'_\ell(k\Delta\eta) = [\partial_\eta\tilde{\Psi}^{(1)}]\,j_\ell(k\Delta\eta) - \partial_\eta[\tilde{\Psi}^{(1)}\,j_\ell(k\Delta\eta)]$.

8

Paper V: Spectro-spatial evolution of the CMB III: transfer functions, power spectra and Fisher forecasts

This paper is the third and final in a series of three papers (see chapter 6 and chapter 7) which together aim to correctly model the frequency domain for the photon phase space distribution (see chapter 2) within the Boltzmann hierarchy (see chapter 3).

With the spectral basis constructed and verifed in the first paper, and the details of perturbation theory explored in the second paper, this paper contains the numerical results for the extended Boltzmann hierarchy. We present photon transfer functions with distorted spectral shapes, CMB power spectra, and forecasts for global energy release.

The distorted SED transfer functions contain three main sources, which allow us to distinguish energy injection times and potentially injection mechanism too. These three sources also manifest in the CMB power spectra which show clearly different spectra for different injection times, including overall sign changes and local acoustic peak structure changes.

We demonstrate that observing cross correlations between distortion amplitudes and temperature anisotropies allows for placing constraints on the total average energy injection. Fisher forecasts reveal that current data from *Planck* will not yield any improvements over *COBE/FIRAS*, however future imagers like LiteBIRD or PICO have potential to improve on current constraints through completely novel means.

Prepared for submission to JCAP

Spectro-spatial evolution of the CMB III: transfer functions, power spectra and Fisher forecasts

Thomas Kite[a] , Andrea Ravenni[b,c] and Jens Chluba[a]

[a]Jodrell Bank Centre for Astrophysics, School of Physics and Astronomy, The University of Manchester, Oxford Road, Manchester, M13 9PL, U.K.

[b]Dipartimento di Fisica e Astronomia "Galileo Galilei", Università degli Studi di Padova, via Marzolo 8, I-35131, Padova, Italy.

[c]INFN, Sezione di Padova, via Marzolo 8, I-35131, Padova, Italy.

E-mail: Thomas.Kite@Manchester.ac.uk, Andrea.Ravenni@unipd.it, Jens.Chluba@Manchester.ac.uk

Abstract. In this paper, we provide the first computations for the distortion transfer functions of the cosmic microwave background (CMB) in the perturbed Universe, following up on paper I and II in this series. We illustrate the physical effects inherent to the solutions, discussing and demonstrating various limiting cases for the perturbed photon spectrum. We clarify the relationship between distortion transfer functions and the photon spectrum itself, providing the machinery that can then compute constrainable CMB signal power spectra including spectral distortions for single energy injection and decaying particle scenarios. Our results show that the $\mu \times T$ and $y \times T$ power spectra reach levels that can be constrained with current and future CMB experiments without violating existing constraints from *COBE/FIRAS*. The amplitude of the cross-correlation signal directly depends on the average distortion level, therefore establishing a novel fundamental link between the state of the primordial plasma from redshift $10^3 \lesssim z \lesssim 3 \times 10^6$ and the frequency-dependent CMB sky. This provides a new method to constrain average early energy release using CMB imagers. As an example we derive constraints on single energy release and decaying particle scenarios. This shows that *LiteBIRD* may be able to improve the energy release limits of *COBE/FIRAS* by up to a factor of $\simeq 2.5$, while *PICO* could tighten the constraints by more than one order of magnitude. The signals considered here could furthermore provide a significant challenge to reaching cosmic variance-limited constraints on primordial non-Gaussianity from distortion anisotropy studies. Our work further highlights the immense potential for a synergistic spectroscopic approach to future CMB measurements and analyses.

Contents

1 Introduction **2**

2 Generalized photon Boltzmann hierarchy **3**
  2.1 Brief recap of the important equations from paper I+II 3
  2.2 Principal sources of anisotropic distortions 5
  2.3 Broad picture for the anisotropic photon spectrum 6
  2.4 Convergence of the photon spectrum 8
  2.5 Change of basis 10

3 Numerical solutions for distortion transfer functions **11**
  3.1 Numerical setup 12
    3.1.1 Switching the physics 12
  3.2 Anisotropies for energy injection in the μ-era 14
  3.3 Anisotropies for energy injection in the residual distortion era 18
  3.4 Anisotropies for energy injection in the y-era 19
  3.5 Anisotropic heating from decaying particles 19
    3.5.1 Time-dependent heating effect on the distortion transfer functions 20
    3.5.2 Perturbed decay effect on the distortion transfer functions 21

4 CMB power spectra with primordial distortions **22**
  4.1 Numerical setup 24
  4.2 CMB temperature power spectrum benchmark 26
  4.3 Single injection CMB power spectra 26
    4.3.1 Isolating various physical effects 27
    4.3.2 Higher-residual power spectra 31
    4.3.3 Distortion auto-power spectra 31
  4.4 Decaying particle CMB power spectra 31

5 Fisher forecasts **33**
  5.1 Accessing information from the residual distortions 40
  5.2 Further optimizing and reducing the observation basis 40
  5.3 Extracting the time of injection 41

6 Discussion and Conclusions **41**

A Perturbed decay term for decaying particle scenarios **50**

B Power spectrum convergence **50**

C Physical effects in SD power spectra **51**

D Correlations with E-modes **51**

E Survey specifications **51**

1 Introduction

The study of perturbations in the primordial plasma has delivered a wealth of cosmological information in the past two decades. Through a combination of theoretical and numerical tools it has been possible to yield not only strong constraints on the initial conditions that seed these perturbations, but also tight limits on the exact constituents of the cosmic inventory [1, 2]. All this insight into the Universe's primordial origins ensued from observations of the photon anisotropies at the last scattering surface, an avenue of discovery in turn made possible by the tight coupling between photons and the rest of the plasma mediated via the baryonic components of the fluid [3–6].

While traditional approaches to studying the early Universe via the Einstein-Boltzmann equations [7–9] capture many aspects of the problem, it is arguable that an entirely novel dimension is still *on the table*. In its complete form, the photon phase space distribution carries dependence on time, spatial coordinates and momentum. Through various manipulations (Fourier transforms and spherical harmonic projections) and assumptions (e.g., Gaussian perturbations) these degrees of freedom are captured with wavenumber k and Legendre moment ℓ. The momentum of the distribution is usually only crudely captured by modelling the frequency spectrum as a blackbody with varying temperature – a consequence of assuming that all energy is thermalised instantaneously in most primordial scenarios. It is well known, however, that the primordial photon spectrum has a greater diversity of spectral shapes at the background level, known as Spectral Distortions (SDs) [10–16].

In the first paper of this series [17, henceforth 'paper I'], we generalised and expanded the traditional average Boltzmann hierarchy to also span the dimension offered through spectral dependence. By understanding the photon frequency hierarchy as a discretised sum over new basis functions, $Y_n(x)$, of dimensionless frequency x, we can accurately model the evolution of the photon spectrum including the residual-era. In the second paper [18, henceforth 'paper II'] we use this discretised formalism to extend the spatial Boltzmann hierarchy, thus completing the triad of variables, leaving no information unexplored in the photon sector of the primordial plasma. This allows us to include the main effects relevant to the evolution of primordial distortion anisotropies, namely, Doppler and potential driving, anisotropic heating, perturbed thermalisation and the full spectral evolution from $y \rightarrow \mu \rightarrow T$ across cosmic history.

We previously showed this method works for the evolution of the background spectrum by replicating the average thermalisation Green's function [19, 20]. In this paper (Sect 2.3), we apply this formalism to the evolution of anisotropic photon spectra. By studying numerical solutions for the spectrum we show that in the presence of average distortions there are three dominant sources of anisotropies. Firstly, and perhaps most familiar, is Doppler boosting of the background spectrum, whether this originates from potential decay or baryonic Doppler driving [21]. The boost operator, $\hat{O}_x = -x\partial_x$, is also responsible for the cosmic microwave background (CMB) temperature and distortion dipole induced by our own motion [22–24]. Secondly we have direct anisotropic heating, where the same mechanism causing a global source of energy will inevitably have some patch-to-patch variations (i.e. via variations in local clocks). Finally there is a source of anisotropies associated with the diffusion of the background spectrum, modulated by local temperature patches [see Eq. (2.1b)].

A crucial step for interpreting perturbed spectra in terms of SED (i.e., spectral energy distribution) amplitudes is discussed in Sect. 2.5. Essentially, there are many ways of describing a spectrum as a series of coefficients, an ambiguity which is important for relating the modelled spectrum and observations (see Sect. 5.2). If one took the SED amplitudes in the Y_n basis at face value they would falsely imply that almost no y and μ distortions are present. In reality because of mutual cancellations and non-trivial overlaps between the modes it is possible to compress the information by *projecting* out the usual SD amplitudes, using only a few residual modes to capture the rest (see paper I). We use this Principle Component Analysis (PCA) technique to show results in a reliable way, which is motivated by the observational procedure.

With these clear definitions of SED amplitude we can calculate transfer functions for different spectral modes (Sect. 3), which thus allows us to present power spectra for the photon spectrum (see Sect 4). This direct link to what would be seen across the CMB sky is a big step in SD cosmology, since we can now infer properties of the background spectrum from the SD anisotropies, and therefore place limits on primordial energy release. Furthermore, we argue it is possible to place limits on the time and details of the injection by studying the shapes and relative heights of the *SD acoustic peaks*. We demonstrate this technique by presenting forecasted constraints on single energy injection and particle decay (Sect. 5). With current data from *Planck* we forecast independent and novel limits which are comparable with *COBE/FIRAS*. With future missions like *LiteBIRD* and *PICO* it is possible to push the limits to be an order of magnitude better than *COBE/FIRAS*, and with potentially much more discriminatory power as to the cause of injection. This opens the exciting opportunity for full spectro-spatial explorations of early-universe and particle physics, bringing CMB anisotropy and spectral distortion science together. In future, this synergy will be further explored and demonstrated, firstly by focusing on the detailed evolution of non-Gaussian perturbations and secondly with detailed forecasts based on *Planck* data using realistic sky simulation.

2 Generalized photon Boltzmann hierarchy

2.1 Brief recap of the important equations from paper I+II

For convenience we briefly summarise the bottom line results from the companion papers, which we refer to for more details and clarification of notation. The treatment of paper I introduces a new set of spectral shapes which in addition to the usual shapes form a sufficiently complete basis to model spectral evolution, as seen by comparing to full binned calculations while reducing the number of equations by at two or three orders of magnitude [25, 26]. In the new formalism, the photon moments are packaged together with SD moments in a vector, $\boldsymbol{y}$, with convention $\boldsymbol{y} = (\Theta, y, y_1, ..., y_n, \mu)^{\rm T}$. These spectral parameters decompose the distortion SED into temperature shift, $G(x)$, y-distortion, $Y(x)$, $n^{\rm th}$ boost of Y, $Y_n(x) = (1/4)^n \hat{O}_x^n Y(x)$, and μ-distortion, $M(x)$. The boost operator is simply $\hat{O}_x = -x\partial_x$ with dimensionless frequency variable $x = h\nu/k_{\rm B}T_z$, where the reference temperature variables scales as $T_z \propto (1+z)$.

The treatment of paper II generalises and extends the standard spatial Boltzmann hierarchy for early-universe perturbations [7–9] to describe the full spectro-spatial evolution of the photon field. As such, many equations remain the same as for the standard Boltzmann hierarchy, unless otherwise stated.[1] Specifically, the gravitational potentials, matter densities and velocities and neutrino perturbations remain unchanged. Spanning the same basis as mentioned above, we define a heating vector $\boldsymbol{Q}$ and thermalisation vector $\boldsymbol{D}$. The former usually only has one non-zero entry contributing to the y-distortion amplitude, while the latter sources Θ from μ to capture the effect of photon production processes. The Kompaneets operator, describing the Compton scattering process in the spectral diffusion problem of the local monopole spectrum, is cast into the same vector space and can thus be represented by a scattering matrix, $M_{\rm K}$, which gradually converts $Y(x)$ to $M(x)$ along a sequence of intermediate $Y_n(x)$ spectra. A similar description exists for the Doppler boosting operator; however, by construction this appears in the equation as $\boldsymbol{b} = M_{\rm B}\boldsymbol{y} + (1, 0, 0, ...)^{\rm T}$, where we have added an inhomogeneous contribution to the Θ component arising from boosts on the background blackbody.

---

[1] In comparison to [9] we use $\Theta_\ell^{\rm Hu} = (2\ell+1)\Theta_\ell$, which also is the definition used in [7]. For Φ and Ψ, we follow the sign convention of [9], which means we have $\Phi^{\rm Ma} = -\Phi$ as defined in [7].

$M_{\rm D} = M_{\rm B}(M_{\rm B} - 3I)$ is similarly associated with the boost operator, being the matrix counterpart of the diffusion operator $\hat{\mathcal{D}}_x = x^{-2}\partial_x x^4 \partial_x$. Both M_K and $M_{\rm B}$ for various representations have been explained in paper I & II and can be found at `www.chluba.de/CosmoTherm`, together with several illustrating videos.

Superscript $X^{(n)}$ shall indicate the order of perturbation, while subscript $X_\ell \equiv \sum_{m=-\ell}^{m=\ell} X_{\ell m} Y_{\ell m}$ indicates the angular moment of the variable[2] (e.g. $\boldsymbol{y}_2^{(1)}$ is the summed quadrupole of the photon vector at first perturbed order). We shall use $\tilde{\boldsymbol{y}}_\ell$ for the corresponding Legendre coefficient. We furthermore use conformal time η to describe the evolution. The photon equations in this extended Boltzmann hierarchy are then given by (see paper II)

$$\frac{\partial \boldsymbol{y}_0^{(0)}}{\partial \eta} = \tau' \theta_z \left[M_{\rm K}\, \boldsymbol{y}_0^{(0)} + \boldsymbol{D}_0^{(0)} \right] + \frac{\boldsymbol{Q}'^{(0)}}{4}, \tag{2.1a}$$

$$\begin{aligned} \frac{\partial \boldsymbol{y}^{(1)}}{\partial \eta} + \hat{\boldsymbol{\gamma}} \cdot \nabla \boldsymbol{y}^{(1)} &= -\boldsymbol{b}_0^{(0)} \left(\frac{\partial \Phi^{(1)}}{\partial \eta} + \hat{\boldsymbol{\gamma}} \cdot \nabla \Psi^{(1)} \right) + \tau' \left[\boldsymbol{y}_0^{(1)} + \frac{1}{10}\boldsymbol{y}_2^{(1)} - \boldsymbol{y}^{(1)} + \beta^{(1)} \chi\, \boldsymbol{b}_0^{(0)} \right] + \frac{\boldsymbol{Q}'^{(1)}}{4} \\ &\quad + \tau' \theta_z \left\{ M_{\rm K}\, \boldsymbol{y}_0^{(1)} + \boldsymbol{D}_0^{(1)} + \left[\delta_{\rm b}^{(1)} + \Psi^{(1)} \right] \left(M_{\rm K}\, \boldsymbol{y}_0^{(0)} + \boldsymbol{D}_0^{(0)} \right) + \Theta_0^{(1)} \left(\boldsymbol{D}_0^{(0)} + M_{\rm D}\, \boldsymbol{y}^{(0)} - \boldsymbol{S}^{(0)} \right) \right\}, \end{aligned} \tag{2.1b}$$

$$\begin{aligned} \boldsymbol{D}^{(0)} &= \left(\gamma_T x_{\rm c}\, \mu^{(0)}, 0, 0, \dots, 0, -\gamma_N x_{\rm c}\, \mu^{(0)} \right)^T, \qquad \boldsymbol{D}^{(1)} = \left(\gamma_T x_{\rm c}\, \mu^{(1)}, 0, 0, \dots, 0, -\gamma_N x_{\rm c}\, \mu^{(1)} \right)^T, \\ \dot{\boldsymbol{Q}}^{(0)} &= \left(0, \frac{\dot{Q}_{\rm c}^{(0)}}{\rho_z}, 0, \dots, 0, 0 \right)^T, \qquad \dot{\boldsymbol{Q}}^{(1)} = \left(0, \frac{\dot{Q}_{\rm c}^{(1)}}{\rho_z} + \Psi^{(1)} \frac{\dot{Q}_{\rm c}^{(0)}}{\rho_z}, 0, \dots, 0, 0 \right)^T, \\ \boldsymbol{S}^{(0)} &= \left(0, \delta_{\gamma,0}^{(0)} + 4\Theta_{\rm e}^{(0)}, -4\Theta_{\rm e}^{(0)}, \dots, 0, 0 \right)^T, \end{aligned}$$

where the first equation describes the effect of energy release on the average CMB spectrum, and the second is for the CMB anisotropies. For details on all terms we refer the reader to paper II, however the most important terms for this paper are discussed in the following paragraphs.

For convenience we also give the Fourier and Legendre transformed form of the equations, where k shall denote the wavenumber of the mode. These equations more closely resemble the traditional implementation of Eq. (2.1) in Einstein-Boltzmann solvers [e.g., 27, 28]

$$\frac{\partial \boldsymbol{y}_0^{(0)}}{\partial \eta} = \tau' \theta_z \left[M_{\rm K}\, \boldsymbol{y}_0^{(0)} + \boldsymbol{D}_0^{(0)} \right] + \frac{\boldsymbol{Q}'^{(0)}}{4}, \tag{2.2a}$$

$$\begin{aligned} \frac{\partial \tilde{\boldsymbol{y}}_0^{(1)}}{\partial \eta} &= -k\, \tilde{\boldsymbol{y}}_1^{(1)} - \frac{\partial \tilde{\Phi}^{(1)}}{\partial \eta} \boldsymbol{b}_0^{(0)} + \frac{\boldsymbol{Q}'^{(1)}}{4} \\ &\quad + \tau' \theta_z \left\{ M_{\rm K}\, \tilde{\boldsymbol{y}}_0^{(1)} + \boldsymbol{D}_0^{(1)} + \left[\tilde{\delta}_{\rm b}^{(1)} + \tilde{\Psi}^{(1)} \right] \left(M_{\rm K}\, \boldsymbol{y}_0^{(0)} + \boldsymbol{D}_0^{(0)} \right) + \tilde{\Theta}_0^{(1)} \left(\boldsymbol{D}_0^{(0)} + M_{\rm D}\, \boldsymbol{y}^{(0)} - \boldsymbol{S}^{(0)} \right) \right\}, \end{aligned} \tag{2.2b}$$

$$\frac{\partial \tilde{\boldsymbol{y}}_1^{(1)}}{\partial \eta} = k \left(\frac{1}{3} \tilde{\boldsymbol{y}}_0 - \frac{2}{3} \tilde{\boldsymbol{y}}_2 \right) + \frac{k}{3} \tilde{\Psi}^{(1)}\, \boldsymbol{b}_0^{(0)} - \tau' \left[\tilde{\boldsymbol{y}}_1^{(1)} - \frac{\tilde{\beta}^{(1)}}{3} \boldsymbol{b}_0^{(0)} \right], \tag{2.2c}$$

$$\frac{\partial \tilde{\boldsymbol{y}}_2^{(1)}}{\partial \eta} = k \left(\frac{2}{5} \tilde{\boldsymbol{y}}_1^{(1)} - \frac{3}{5} \tilde{\boldsymbol{y}}_3^{(1)} \right) - \frac{9}{10}\, \tau'\, \tilde{\boldsymbol{y}}_2^{(1)}, \tag{2.2d}$$

$$\frac{\partial \tilde{\boldsymbol{y}}_{\ell \geq 3}^{(1)}}{\partial \eta} = k \left(\frac{\ell}{2\ell + 1} \tilde{\boldsymbol{y}}_{\ell-1} - \frac{\ell + 1}{2\ell + 1} \tilde{\boldsymbol{y}}_{\ell+1} \right) - \tau' \tilde{\boldsymbol{y}}_\ell^{(1)}, \tag{2.2e}$$

---

[2] In rare cases of denoting the angular moment of one of the $Y_n(x)$ amplitudes we will use $y_{n,\ell}$ as the convention, and similarly will rename the y-distortion amplitude $y_{0,\ell}$. In many cases we simply label the ℓ explicitly for clarity.

and can be solved using stiff ordinary differential equation (ODE) routines [29]. The equation set takes a form that is extremely similar to the standard photon brightness temperature equation with the differences that i) the average CMB monopole can evolve, ii) Doppler and potential driving terms now affect various spectral parameters and iii) the local monopole sees new effects from thermalisation process and energy injection. Some first discussion of the expected physical effects was already given in paper II. Here, we will now demonstrate all these using numerical solutions of the transfer functions, and illustrate how they eventually affect the CMB signal power spectra. Note that we have not included polarisation effects in our description of the spectro-spatial problem; however, this should not affect the main conclusions significantly. We have included polarisation effects for the standard Θ, which on the one hand allows us to compare power spectrum solutions with `CLASS`, and on the other hand provides important cross correlations between SD and E-modes (see Sect. 5).

2.2 Principal sources of anisotropic distortions

It will be useful for interpreting the following sections results to pause and discuss some features of Eq. (2.1). Firstly we note the presence of distinct timescales: Thompson terms are weighted by τ' while Kompaneets and thermalisation terms are weighted by $\tau'\theta_z$, where $\theta_z = k_{\rm B}T_z/m_{\rm e}c^2$ is the dimensionless temperature variable. This implies the former is the dominant interaction, however only affecting higher multipoles of the distribution, leaving the latter as the dominant term for the monopole. We note that $\tau'\theta_z$ decreases with time, lending it greater importance in the μ-era. Furthermore the production of photons carries an implicit timescale in the critical frequency $x_{\rm c}$, effectively shutting off photon creation for $z \lesssim 2\times10^5$ [30].

Following some mechanism of average energy injection $\boldsymbol{Q}'^{(0)}$ which forms the background distortion, three main sources of anisotropic distortions are present: boosting, anisotropic heating and perturbed thermalisation.

- Firstly the boosted background spectrum $\boldsymbol{b}_0^{(0)}$ appears twice in Eq. (2.1b), once as gravitational boosting which is strongly associated with horizon crossing, and again as the Doppler boosting from local baryon velocities $\beta^{(1)}$. This simply sources the boosted spectrum, e.g. for early energy injection times there will be a spectrum resembling $\hat{O}_x M(x)$ sourced in local patches. One hallmark of the boosting effect is that early time injection yields $\hat{O}_x M(x)$, which gives a same-sign combination of $y^{(1)}$ and $\mu^{(1)}$ from performing the PCA projection. On the contrary late time injection yields $\hat{O}_x Y(x)$, which gives an opposite-sign combination of $y^{(1)}$ and $\mu^{(1)}$.

- The second source is from direct anisotropic heating, which can be from modulations of the background heating $\simeq \Psi^{(1)} Q_{\rm c}'^{(0)}$ or from an explicit model dependent heating term $Q_{\rm c}'^{(1)}$ (below we will consider the heating from decaying particles which is thus modulated by $\delta_{\rm dm}$). These terms arise momentarily from energy injection, and then undergo thermalisation through $M_{\rm K}\boldsymbol{y}_0^{(1)} + \boldsymbol{D}_0^{(1)}$. There is one more term following this behaviour other than these two explicit ones: while arising from Kompaneets scattering, the term $\Theta_0^{(1)}\dot{Q}_{\rm c}^{(0)}/4\tau'\theta_z = \Theta_0^{(1)}\left(\Theta_{\rm e} - \Theta_{\rm eq}\right) \in \Theta_0^{(1)}\boldsymbol{S}$ in practice resembles a modulation to heating. It arises from terms associated with electron heating, and importantly carries the inverse time scale $\tau'\theta_z$ which makes it manifest at late times unlike other scattering terms. Physically this is because at early times the electrons quickly reaches equilibrium with photons ($\Theta_{\rm e} \approx \Theta_{\rm eq}$). At late times however we see this term change the details of energy injection to the local photon patch. To avoid the risk of introducing a misnomer we clarify: this term does not inject energy, but simply changes which spectral shape is excited, with a shift between $Y(x)$ and $Y_1(x)$.

- The third and final source is perturbed thermalisation, including perturbed scattering effects $\propto \left[\delta_\mathrm{b}^{(1)} + \Psi^{(1)}\right] M_\mathrm{K} \boldsymbol{y}_0^{(0)}$ and perturbed emission $\propto \left[\delta_\mathrm{b}^{(1)} + \Psi^{(1)} + \Theta_0^{(1)}\right] \boldsymbol{D}_0^{(0)}$. These simply modify the local thermalisation timescale of $M_\mathrm{K}\boldsymbol{y}_0^{(1)} + \boldsymbol{D}_0^{(1)}$ according to the average spectrum. Also within perturbed scattering we find $M_\mathrm{D}\boldsymbol{y}^{(0)} - \boldsymbol{S}^{(0)}$, once the aforementioned heating term has been extracted. This part of perturbed scattering sources a local spectral shape resembling the diffusion operator applied to the background together with a shift from Y to Y_1 according to Θ_eq. All of these effects are typically important at earlier times only.

These three sources are shown in Fig. 1. Furthermore see Sect. 3.1.1 for more details on these *physics switches* which we make extensive use of to distil the physical picture throughout the paper.

We provide a disclaimer for the choice of groupings both for sources and for the switches introduced later: there is no unique choice of this decomposition, and many terms fit into multiple categories from a physical point of view. Consider for example the heating term which has been extracted from $\boldsymbol{S}^{(0)}$ whose origin is in the physics of Kompaneets scattering, however its behaviour can be thought of as a form of anisotropic heating. Even the term $\Psi^{(1)} Q_\mathrm{c}'$ is associated with local thermalisation efficiency, and is not a direct form of energy injection per se. Generally, considering this paper is largely concerned with the presentation of numerical results, we have taken a qualitative view of bottom line behaviour rather than a fundamental view of the underlying physics when choosing our grouping of terms.

2.3 Broad picture for the anisotropic photon spectrum

In paper I, we showed how the extended Y_n frequency basis can accurately capture the evolution of the background spectrum by reproducing much more expensive binned frequency calculations within `CosmoTherm` [19, 31]. The goal in this paper is to apply this new basis to the generalised Boltzmann hierarchy and explore the evolution of the anisotropic photon spectrum. In contrast to the background spectrum, this will depend on wavenumber k and angular scale ℓ, as is familiar from usual early-universe perturbation theory, a fact which renders the usual binned spectral treatments prohibitively expensive.

To study the three main anisotropic distortion sources we numerically solve scenarios with $\Delta\rho/\rho = 10^{-5}$ Dirac-δ energy injection at fiducial times $z = 5 \times 10^5$ (μ-era), $z = 5 \times 10^4$ (residual-era), $z = 5 \times 10^3$ (y-era). In Fig. 1 we show the corresponding spectrum for $k = 0.01\,\mathrm{Mpc}^{-1}$ at three time slices, showing the time dependence of different sources. Note that we show the purely distorted spectrum with energy dimensions $x^3 \Delta f(x)$, meaning we have subtracted the local temperature shift $\Theta^{(1)}$. This is common throughout the paper to avoid inflationary perturbations dominating the figures (typically $\simeq 10^5$ larger).

Typically speaking the leftmost panels will show the spectrum shortly after the energy injection (corresponding to super-horizon state for the top two rows). The rightmost panel shows the spectrum at late times – around recombination or later. We see that in all cases the boosting sources grow strongly form left to right, starting at horizon crossing (gravitational boosting) and continuing sub-horizon (baryonic Doppler boosting). The other two sources are only important for early injection times, and dominate over the boosting sources deep in the μ-era. Notice that the earliest injection times yield anisotropic spectra with unfamiliar three-peak structure arising from both $M_\mathrm{D}\boldsymbol{y}^{(0)}$ and $Y_1 - Y$ (see perturbed thermalisation term), and cannot be easily recognised as a simple y or μ spectrum. Anisotropic heating on the other hand initially sources y (with a small y_1 correction), which then has the opportunity thermalise via the equivalent terms to the average spectrum [see first terms in second row of Eq. (2.1b) in comparison to Eq. (2.1a)], and thus follows the same evolution as the average distortion picture. For example, the spectra from anisotropic heating cross the zero at $x \simeq 1$ and $x \simeq 2$

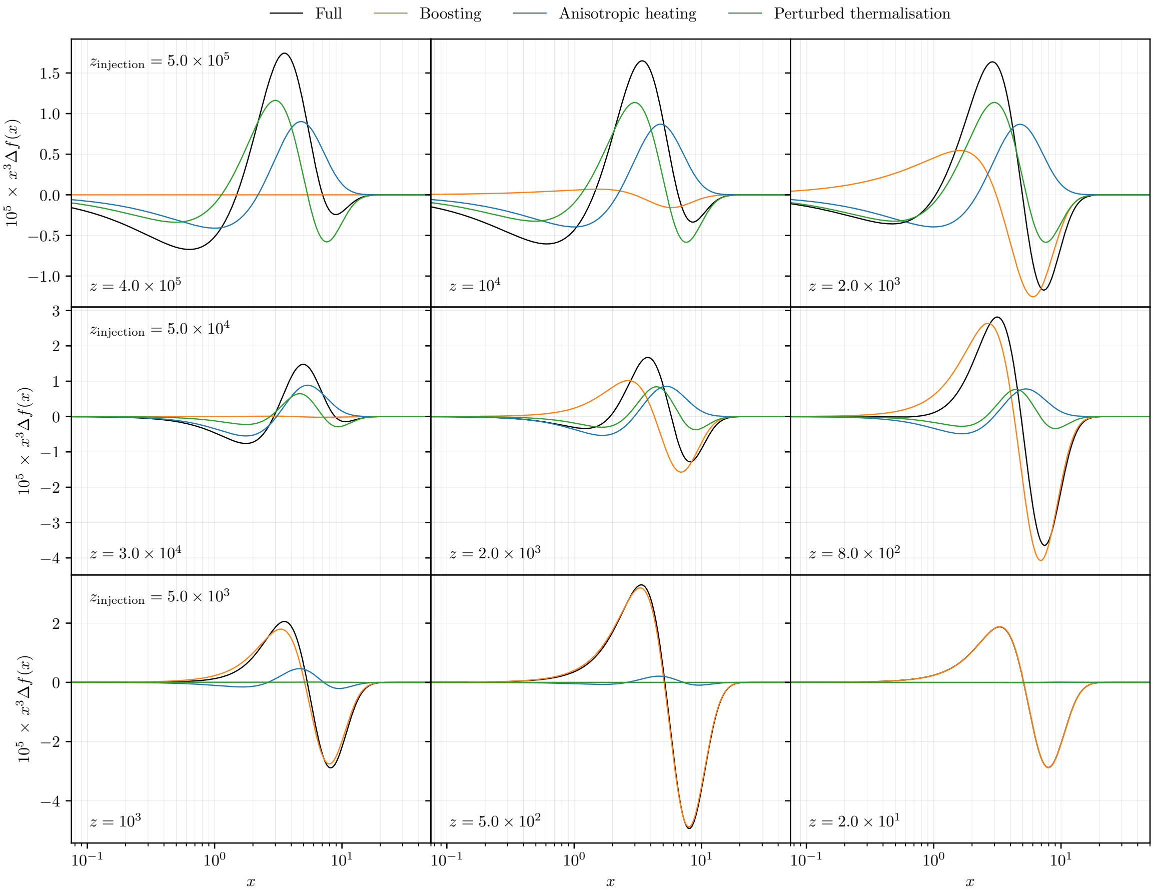


Figure 1: Snapshots of the SD spectrum for $\ell = 0$ and $k = 0.01\,\text{Mpc}^{-1}$ arising from several injection redshifts. The rows show, from top to bottom, injection at $z = 5 \times 10^5$, $z = 5 \times 10^4$ and $z = 5 \times 10^3$. Left to right show different stages of evolution for each injection scenario. Coloured lines in each panel show the spectrum with only one class of source terms included.

for μ and residual-era injection respectively, emulating the usual three era picture. The anisotropic heating spectrum in the third row however does not correspond simply to a y distortion due to the term $\propto \Theta_0^{(1)}\,[Y_1(x) - Y(x)] \in \Theta_0^{(1)}\boldsymbol{S}^{(0)}$, which has no opportunity to thermalise in the late Universe.

All of these spectral shapes can be recognised in Fig. 2, where each of the important operators are demonstrated. For example, we can see that spectral shapes with the three peak structure arise from $\hat{\mathcal{D}}_x$ (early injection perturbed thermalisation) and $Y_1 - Y$ (late time anisotropic heating). In the middle panel we show boosted spectra, where it is important to note $\hat{O}_x M$ projects onto a same-sign mix of Y and M and is well captured by these two numbers. On the other hand, $\hat{O}_x Y$ gives an opposite-sign mix, and additionally needs around two residual modes to converge (see Sect. 2.5). This dependence on residual modes will manifest later in late time injection power spectra (see Fig. 21). Likewise, $\hat{\mathcal{D}}_x M$ requires at least two residual modes to converge, however, we will see later that early time injection power spectra do not in fact make as heavy use of residual mode information, likely from some cancellation of residual modes with other sources. Comparing individual SED amplitudes here can be somewhat misleading considering the different energies they carry (this leads to μ often being $\approx 1.401/0.25 = 5.6$ times larger than y), but we can loosely assert that boosting the average y distortion – the dominant late time behaviour – yields a μ amplitude four times the size of y, however

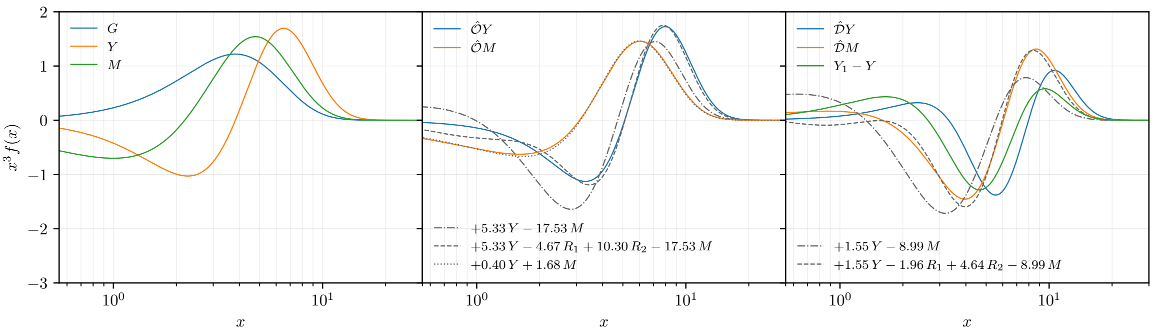


Figure 2: A figure showing the main types of spectral shape. Note that they have been re-scaled by arbitrary constants to make them all comparable. The leftmost panel shows the usual shapes familiar from average frequency evolution, and also the limiting case for anisotropic injection before the y-era. The middle panel shows the boosted spectra for both Y and M. The rightmost panel shows the diffusion operator acting on Y and M as well as $Y_1 - Y$, shapes that all emerge from early time thermalisation effects and late time anisotropic heating (recognisable by their three peak structure). Some numerical approximations are shown in gray lines.

$\hat{\mathcal{D}}_x M$ carries closer to six times as much μ as y. This will be further exacerbated by anisotropic heating thermalising to a μ distortion. We will verify later that the early universe injection will yield much stronger $\mu \times \mu$ correlations than $y \times y$ (see Fig. 22).

2.4 Convergence of the photon spectrum

When studying the background spectrum there is the luxury of comparing to the full `CosmoTherm` calculation (see Sect. 2.8 in the companion paper I). This has demonstrated that the ODE representation of the thermalisation problem is highly accurate and captures the main physical features of the full treatment. For anisotropies, however, the parameter space grows in many dimensions, and a direct `CosmoTherm` convergence benchmarks would be quite expensive. Despite this limitation we can expect the anisotropic treatment to perform well. It can be seen in Eq. (2.1) that the sources of anisotropies arise either from direct sources of y, or the matrix forms of the boost and diffusion operators ($\hat{O}_x \to M_{\rm B}$, $\hat{\mathcal{D}}_x \to M_{\rm D}$), precisely the operators around which the spectral basis was constructed. The bottom line is that the process of boosting is captured exactly in this formalism, not approximately, as long as the dominant part of a spectrum is relying mostly on $N < N_{\rm max}$. More concretely, the only boosted SEDs not directly contained in the basis are $\hat{O}_x Y_{N_{\rm max}}(x)$ and $\hat{O}_x M(x)$. For the first case we note that $y_{N_{\rm max}}$ becomes smaller for growing $N_{\rm max}$, with y_{15} only seeing significant contributions in very narrow windows of the residual-era ($z \approx 10^5$ providing a worst case scenario). For the second case we have shown $\hat{O}_x M(x)$ to map extremely well back into the basis even with $N_{\rm max} \leq 1$ (see paper I). This all means that the accuracy of the anisotropic evolution will typically be limited by the accuracy of the average evolution.

In Fig. 3 we show the photon spectrum for $0 \leq N_{\rm max} \leq 15$ at various single energy injection redshifts (from top to bottom row: $z_{\rm injection} = 5 \times 10^5,\ 10^5,\ 5 \times 10^4,\ 5 \times 10^3$). From left to right we show different stages of the evolution, from the stable super-horizon state through to late post-recombination evolution. We can immediately see that earlier injection times see poorer convergence than later times, with $z = 10^5$ performing the worst as expected. This is due to the only late time source being the boost operator, which given the arguments above is well captured in this basis. The greater diversity of sources for the μ-era injection puts more strain on the numerical method, however, we note that between $N_{\rm max} = 15$ and $N_{\rm max} = 13$ we only see sub-percent changes in the right most panels of Fig. 3. This statement depends on the moment you observe the spectrum, which is why we

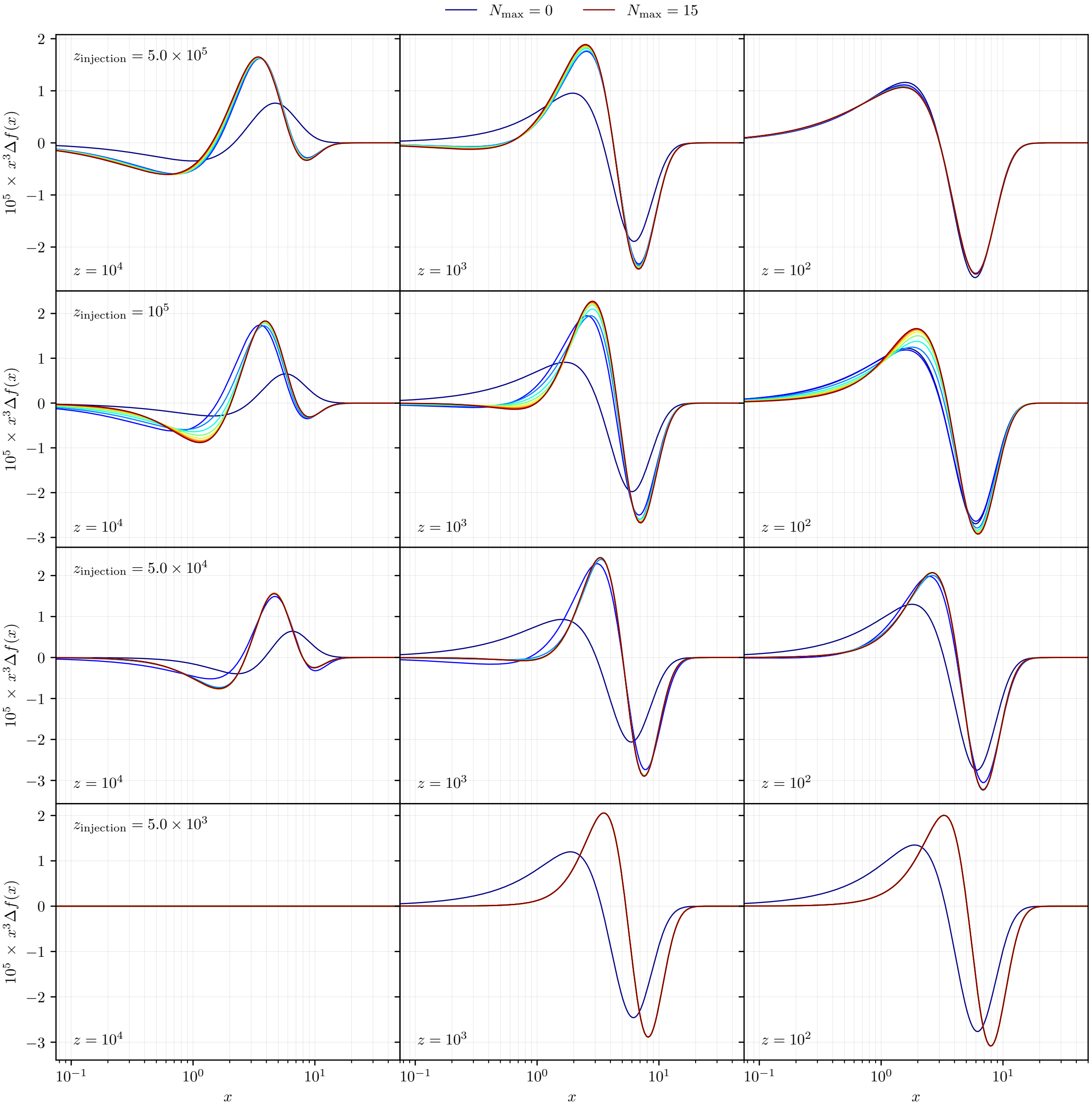


Figure 3: Snapshots of the SD spectrum for $\ell = 0$ and $k = 0.01\,\mathrm{Mpc}^{-1}$ arising from several injection redshifts. Within each panel, we show the solution when varying the size of the computation basis. For $z_{\text{injection}} \lesssim 5 \times 10^3$ and $z_{\text{injection}} \gtrsim 5 \times 10^5$ convergence is extremely rapid, while for intermediate cases, a basis with Y_{15} starts to shows its limitations.

opted to study the spectra in Fig. 3 at the same moments of time in a given column, unlike Fig. 1 where we prioritised elucidating physical sources at various moments of evolution. We note that in Fig. 32 and Fig. 33 we perform a similar convergence analysis for the power spectra which gives a less time dependent sense of the performance of the basis.

Notably the spectrum for y-era injection (bottom row) is captured almost exactly even with just Y_1 since the expected limiting case $\hat{O}_x Y = 4Y_1$ is precisely captured in that basis (late energy injection sees very little contribution from other sources, see Fig. 1). The μ-era injection (top row of Fig. 3) shows similarly good result. The main source in this era is less clear than for late times, but is

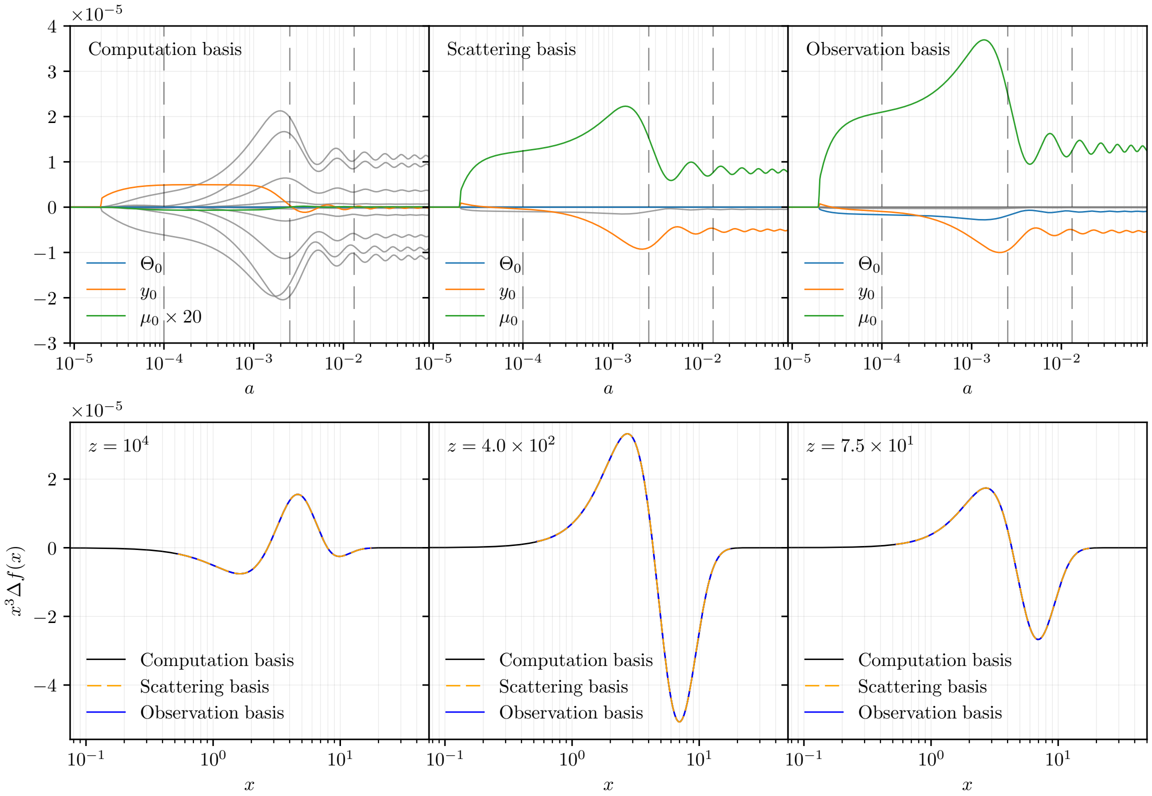


Figure 4: Differences between monopole SED transfer functions in various bases (upper panels) and the corresponding SD signals at three snapshots (lower panels). The transfer functions correspond to waves with $k = 0.01\,\mathrm{Mpc}^{-1}$ following an energy injection of $\Delta\rho/\rho = 10^{-5}$ at $z = 5 \times 10^4$. The amplitude of Θ is subtracted prior to changing basis, to avoid hiding the signal under primordial fluctuations five orders of magnitude larger. In this way, we highlight the fictitious Θ from the unconstrained change of basis (observation basis). The solid/gray lines in each panel represent the residual modes for the given basis. In the computation basis, these higher order contributions exceed the y and μ transfer functions, while for the scattering and observation basis, they contribute marginally, as anticipated. In spite of the drastic SED parameter differences for the three basis, the basis-independence of the SD signal is demonstrated in the lower panels.

some mix of $\hat{O}_x M(x)$, $\hat{\mathcal{D}}_x M(x)$, $M(x)$ and $Y_1(x) - Y(x)$, all of which are captured well in this basis. It should be highlighted that the transition from T to μ-era is slower in this formalism compared to usual numerical solutions and thus a study of convergence here is not the whole picture. In essence some small thermalisation from $\mu^{(0)}$ to $\Theta^{(0)}$ occurred where we would not have expected any; however, the correction is small and can likely be eliminated with further improvements of the thermalisation treatment (see paper I for discussion).

2.5 Change of basis

It is explained in paper I that there is a large degree of degeneracy between the $Y_n(x)$ spectral shapes and the more common $G(x)$, $Y(x)$ and $M(x)$. Upon solving the average spectral evolution this led to seemingly very different energy branching ratios for SED amplitudes that otherwise converged to the same expected resulting photon spectrum. This problematic disconnect between branching ratios (or soon transfer functions) can be remedied by performing a *change of basis* in which a new set of basis function SEDs are chosen that better suit the physics being studied.

In particular one can *prioritise* the regular SED shapes by first performing a projection onto $G(x)$, $Y(x)$ and $M(x)$, and subsequently constructing a set of orthogonal SED through PCA. We will construct two such bases: firstly the *observation basis*, which as the name suggests performs a PCA on the binned frequency space[3] to allow direct comparison to results that are akin to what would be obtained in real measurements. This basis was first considered in [32] and compresses the accessible signal information significantly. Secondly, the *scattering basis*, which is constructed as before, but with the additional constraint of photon number, so no other SED will project to $G(x)$ and vice-versa. This final basis highlights the fact that subtracting the theoretically accurate $G(x)$ from an observationally acquired spectrum is usually not possible provided the finite binning of an experiment. In this light the $Y_n(x)$ basis will be called the *computation basis*, since it has been constructed to model the boosting/diffusion/scattering processes dictating the evolution of photons in plasma. This new vocabulary emphasises the *basis-independence* of the CMB spectrum anisotropies in contrast to the *basis-dependence* of SED amplitudes – a fact which should be present in the readers mind when interpreting any results in the subsequent sections in this paper, especially for detection prospects (Sect. 5).

While transfer functions will be discussed extensively in Sect. 3, we show a single example in Fig. 4 to illustrate the difference in basis choice. Focusing on the upper row first, the upper left panel shows the computation basis with residual modes $y^{(1)}_{n,0}$ typically dominating over the $y^{(1)}_{\ell=0}$ and $\mu^{(1)}_{\ell=0}$ amplitudes. This is understood since the background spectrum consists of a wide mix of $y^{(0)}_n$ following energy injection in the residual-era, as these get boosted to perturbed spectra with $y^{(1)}_{n+1,1}$. Only for $y^{(0)}_{N_{\rm max}}$ and $\mu^{(0)}$ does this boosting mix directly into $y^{(1)}_\ell$ or $\mu^{(1)}_\ell$. The upper middle panel in the upper row of Fig. 4 shows the results of casting to the scattering basis – projecting the spectrum back onto the main SED and using residual modes to capture the remaining signal. This can indeed be seen to give the same spectral shape while compressing the information to the usual SD amplitudes and just a small contribution from residual modes (see lower panels). Finally the upper right panel shows the observation basis, representing what could be seen with a binned observation of the sky. Most notably it can be seen that some $\Theta^{(1)}_\ell$ is generated by counteracting an increase in $\mu^{(1)}_\ell$, a result of inferring spectral shapes from a limited window of visibility.[4] However, the representation of the signal is independent of these parametrisation aspects (lower panels in Fig. 4).

3 Numerical solutions for distortion transfer functions

With some understanding of SED transfer functions and how these map to a corresponding distorted photon spectrum, we are now in the position to gain a more intuitive understanding of the behavior of distortion anisotropies and their evolution. Like for the thermalisation Green's function it is instructive to first consider single redshift injections of average energy from which we will then distill some of the physics for distortion modes at various scales. All results in the following section are shown in the scattering basis (see Fig. 4 for illustration), meaning that μ and y are usually representative of the spectrum, with only minor contributions from the residual modes which will not be highlighted here. The primordial temperature fluctuations have not been subtracted this time, allowing for comparison of relative phases between the SED parameter amplitudes.

---

[3] We take a fiducial binning from $\nu_{\rm min} = 30\,$GHz to $\nu_{\rm max} = 1000\,$GHz with steps $\Delta\nu = 1\,$GHz

[4] This effect is familiar in different settings [19, 20].

3.1 Numerical setup

To solve the coupled system of Boltzmann equations we extend the anisotropy module of `CosmoTherm` [25]. We set adiabatic initial conditions for the standard perturbations while the distortion parameters are initially set to zero, given that no initial inflationary distortion signals are expected. The ODE system is solved using a sixth order Gear's method with adaptive time-stepping. This method is stiffly-stable and does not require any separate treatment in the tight-coupling regime. The corresponding solver was implemented to solve the cosmological recombination problem [29, 33]. A relative precision of $\simeq 10^{-4}$ is requested and redshift is used as the main time-variable.

We truncated the multipole hierarchy following [7]. Depending on the scale, we include a varying number of multipoles for CMB temperature and polarisation anisotropies, neutrinos and the distortion parameters. We find that $\ell_{\rm max} = 15$ is sufficient to achieve accurate power spectrum results; however, for the transfer functions in this section we expand this greatly (up to $\ell_{\rm max} = 100$) to ensure no *reflected* energy in the shown time intervals. We do not include reionisation or perturbed recombination effects in our treatment. Also, polarisation effects are only treated carefully for the temperature perturbations, not for the distortion parameters. However, these approximations are not expected to change the overall picture significantly.

3.1.1 Switching the physics

For the results presented below, it is instructive to switch on/off various physical effects. The goal is to illustrate the effect on the distortion anisotropies, so in all cases, we do not modify the standard perturbation equations for temperature and polarisation terms. We introduce various physical switches in relation to the sources mentioned in Sect. 2.2 (also see Fig. 1): Doppler/potential boosting, perturbed emission/scattering, and anisotropic heating.

- Referring to Eq. (2.2), switching off **Doppler** boosting (here sometimes also referred to as Doppler driving) means we drop the term $\propto \tilde{\beta}^{(1)}\boldsymbol{b}_0^{(0)}/3$ in the dipole equations of the distortions.

- Similarly, to switch off **potential** driving we drop the terms $-\partial_\eta \tilde{\Phi}^{(1)}\boldsymbol{b}_0^{(0)}$ and $k\tilde{\Psi}^{(1)}\,\boldsymbol{b}_0^{(0)}/3$ in the monopole and dipole equations of the distortions. These two switches are presented together simply as **boosting**.

- Perturbed **emission** off means not accounting for the group of terms $\propto \left[\tilde{\delta}_{\rm b}^{(1)} + \tilde{\Psi}^{(1)} + \tilde{\Theta}_0^{(1)}\right]\boldsymbol{D}_0^{(0)}$ in the monopole distortion equation.

- Similarly perturbed **scattering** off means not accounting for the group of terms

  $$\propto \left[\tilde{\delta}_{\rm b}^{(1)} + \tilde{\Psi}^{(1)}\right] M_{\rm K}\,\boldsymbol{y}_0^{(0)} + \tilde{\Theta}_0^{(1)}\left(M_{\rm D}\boldsymbol{y}_0^{(0)} - \boldsymbol{S}^{(0)}\right)$$

  in the monopole distortion equation, *except* the aforementioned terms within $\boldsymbol{S}^{(0)}$ which are deemed anisotropic heating (see Sect. 2.2). These previous two switches together make up **perturbed thermalisation**.

- Finally neglecting **anisotropic heating** means omitting all terms within $\boldsymbol{Q}'^{(1)}/4$, and also the terms $\propto Q_{\rm c}'/4\tau'\theta_z$ within $\boldsymbol{S}^{(0)}$.

The purely spatial thermalisation terms $M_{\rm K}\,\tilde{\boldsymbol{y}}_0^{(1)} + \boldsymbol{D}_0^{(1)}$ are always switched on, meaning there is a similar evolution for spatial spectra as for average spectra. This is not to say that all sources will undergo a simple thermalisation process, since terms like $M_{\rm D}\,\boldsymbol{y}^{(0)}$ will continuously source from

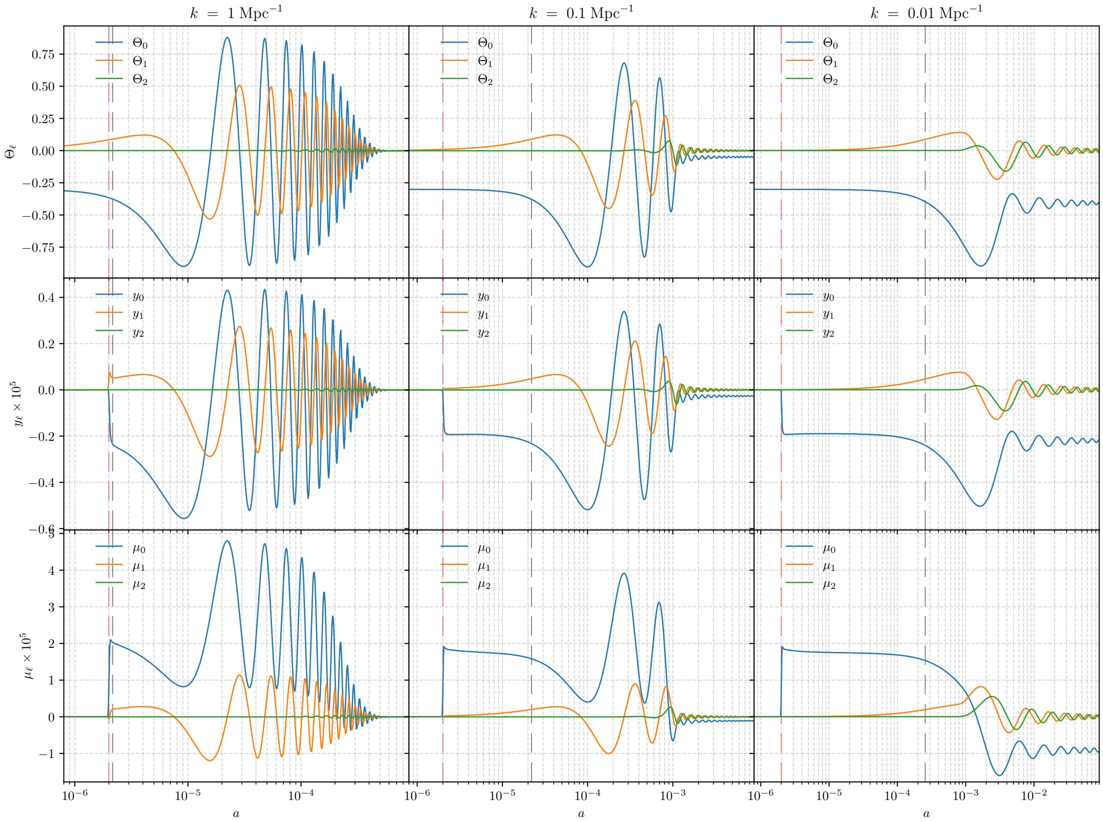


Figure 5: CMB transfer functions for single heating at $z_{\rm injection} = 5{\times}10^5$ with $\Delta\rho_\gamma/\rho_\gamma = 10^{-5}$ and wavenumber k as labeled. This leads to an average distortion that freezes with $\Theta^{(0)} \approx 4.9{\times}10^{-6}$, $y^{(0)} \approx 0.0$ and $\mu^{(0)} \approx 1.1{\times}10^{-5}$ at $z \lesssim 10^5$. The numerical solutions are computed using $N_{\rm max} = 15$ with a rotation to the scattering basis (Sect. 2.5). Dashed vertical lines show times of horizon crossing (gray) and energy injection (red).

the average spectrum, and boosting sources typically only occur after the early-time thermalisation window (this is true for k modes which influence the CMB power spectrum).

One small clarification about the nomenclature of the perturbed thermalisation terms is in order. Physically, the thermalisation process requires the combined action of Compton scattering and DC/BR emission and absorption [11, 14, 16, 34]. When we say ‘perturbed scattering’, we mean ‘perturbed Compton scattering’ as opposed to ‘perturbed Thomson scattering’, which has no effect on the spectral shape but would only slightly modify the Thomson visibility function, leading to a higher order effect [e.g., 35]. The term ‘perturbed emission’ is indeed somewhat misleading as it includes the change in the balance between DC/BR emission and Compton up-scattering, which ultimately defines the distortion visibility [25, 30]. This latter effect was estimated by [36] in the context of primordial non-Gaussianity, and originates from changes in the thermalisation efficiency around $z_\mu \simeq 2 \times 10^6$ due to the presence of perturbations. To not confuse it with ‘perturbed thermalisation’ (which includes all terms), we shall choose to use ‘perturbed emission’ instead of e.g., ‘perturbed thermalisation efficiency’ or ‘perturbed visibility’.

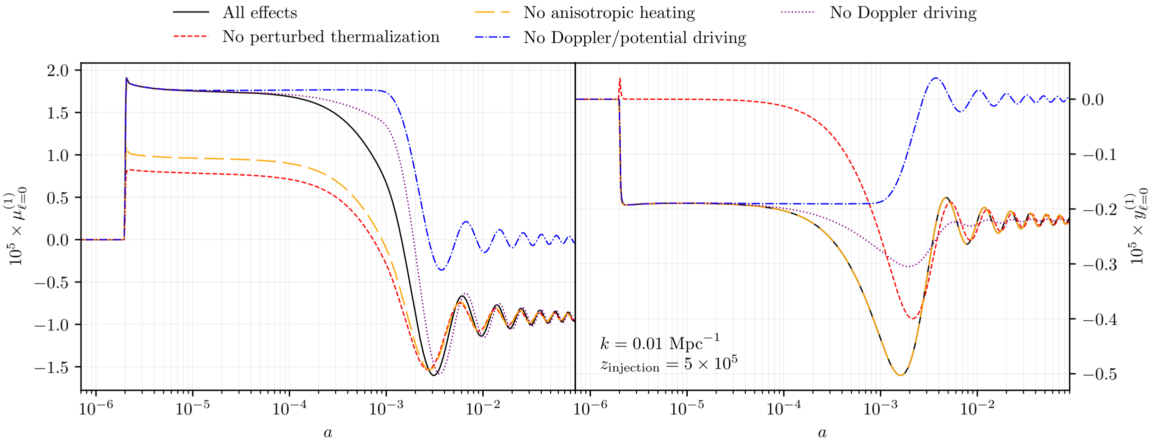


Figure 6: Same as Fig. 5, but focusing on $\mu_0^{(1)}$ (left column) and $y_0^{(1)}$ (right column) for $k = 0.01\,\text{Mpc}^{-1}$. Perturbed thermalisation effects (here dominated by DC and BR emission effects) are switched off for the red/dashed line, showing a reduction of around half the initial local distortion, however illustrating that the overall late evolution is not affected significantly by these corrections. Excluding the effect of Doppler and potential driving for the distortion anisotropies shows that without potential driving the modes simply decay and oscillate around a zero point. Doppler driving adds a small correction to this picture, most important around recombination. Switching off perturbed thermalisation leads the local y distortion to only show an initial transient spike, which rapidly thermalises through $M_{\rm K}\boldsymbol{y}_0^{(1)} + \boldsymbol{D}_0^{(1)}$ (never switched off). Anisotropic heating (orange) contributes about 1/2 of the initial $\mu^{(1)}$ distortion, and does not source $y^{(1)}$. Finally, only neglecting Doppler driving shows clearly how the slow part of the evolution is dominated by potential decay (purple/dash-double-dotted line).

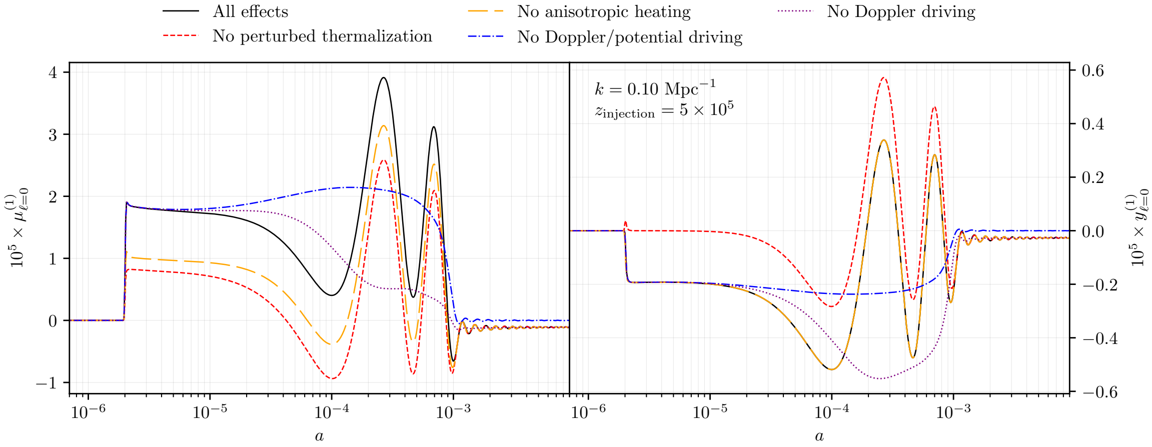


Figure 7: Same as Fig. 6, but for $k = 0.1$ Mpc^{-1}.

3.2 Anisotropies for energy injection in the μ-era

We begin our analysis by considering average energy release deep into the μ-era ($z > 3 \times 10^5$). We illustrate the transfer functions for $\Theta^{(1)}$, $y^{(1)}$ and $\mu^{(1)}$ in the various columns of Fig. 5 for a single heating occurring at redshift $z_{\rm injection} = 5 \times 10^5$, varying the wavenumber of the mode in the columns. In this figure (as well as Fig. 8 and Fig. 11) the transfer functions for $\Theta^{(1)}$ behave all as expected and well-known for adiabatic perturbations [e.g., 6]. Similarly, as expected, distortion anisotropies

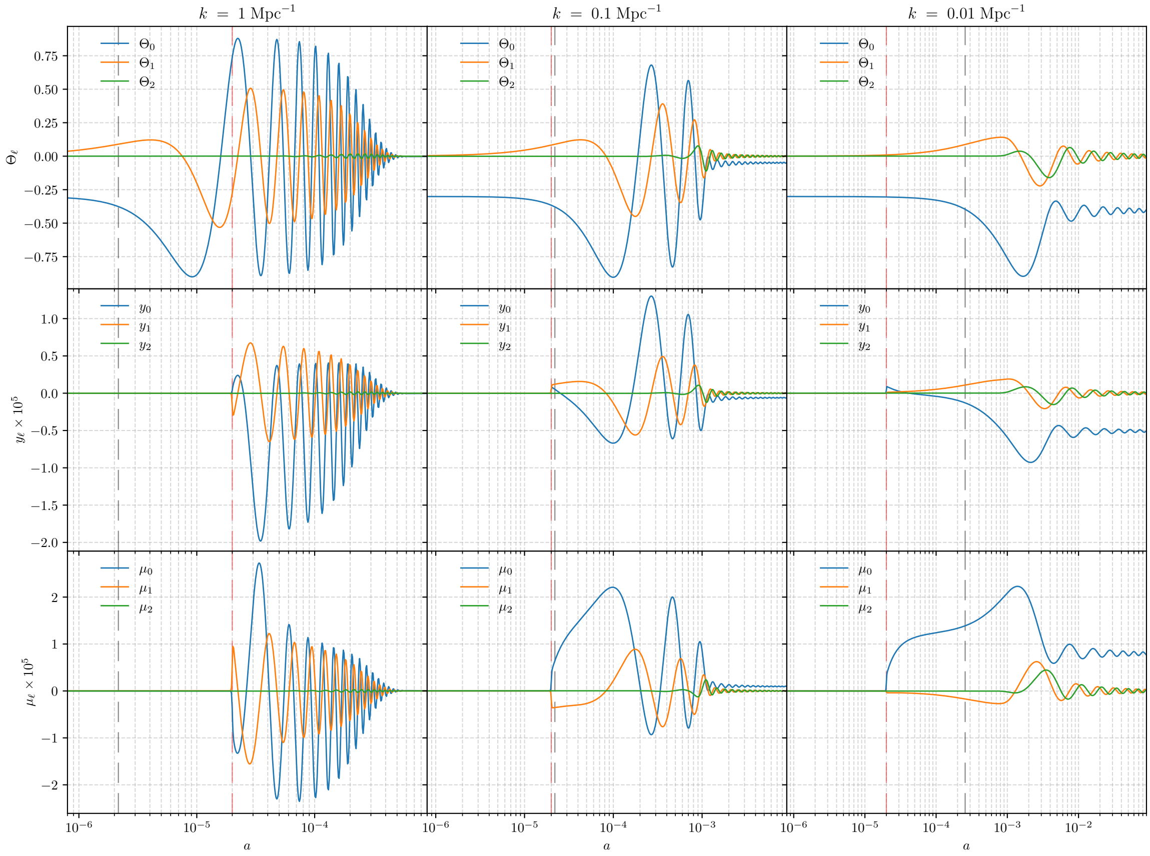


Figure 8: CMB transfer functions for single heating at $z_{\text{injection}} = 5 \times 10^4$ with $\Delta\rho_\gamma/\rho_\gamma = 10^{-5}$ and wavenumber k as labeled. This leads to an average distortion that freezes with $\Theta^{(0)} = 7.5 \times 10^{-9}$, $y^{(0)} = 3.9 \times 10^{-7}$ and $\mu^{(0)} = 1.2 \times 10^{-5}$ at $z \lesssim 8000$. The numerical solutions are computed using $N_{\max} = 15$ with a rotation to the scattering basis. Dashed vertical lines show times of horizon crossing (gray) and energy injection (red).

only become visible after the average distortion is present. Additionally Fig. 6 and Fig. 7 show the $k = 0.01\ \mathrm{Mpc}^{-1}$ and $k = 0.1\ \mathrm{Mpc}^{-1}$ modes with various physics switches on/off.

Focusing on the early evolution we see that following the creation of an average distortion there is both a local monopole y and μ sourced. We can see by inspecting the upper-left panel of Fig. 1 that this is equal parts from anisotropic heating and the perturbed scattering terms, as verified in Fig. 6. This evolution quickly reaches an equilibrium state, where the mode then waits till horizon crossing, upon which the boosting effects from gravitational potential decay and doppler boosting begin. These negatively drive both μ and y, where the equal sign is characteristic of $\hat{O}_x M(x)$ [we will see opposite sign mixes later from $\hat{O}_x Y(x)$]. At late times the distortion SED transfer functions oscillate around a varying mean, mostly driven by the gravitational potentials, with small corrections from baryonic Doppler boosts, again as seen from Fig. 6. Note, however, that the time of recombination receives large contributions from baryonic Doppler driving (red line in Fig. 6), which makes it an important source to CMB power spectra. By further inspecting Fig. 7 we can distinguish that the oscillations in the tight coupling phase are associated with potential driving, while the baryonic Doppler boosts mainly contribute at horizon crossing.

There is a small transient phase of evolution before reaching the superhorizon equilibrium state (also seen well in the right panel of Fig. 6). This can be seen as the equivalent thermalisation process

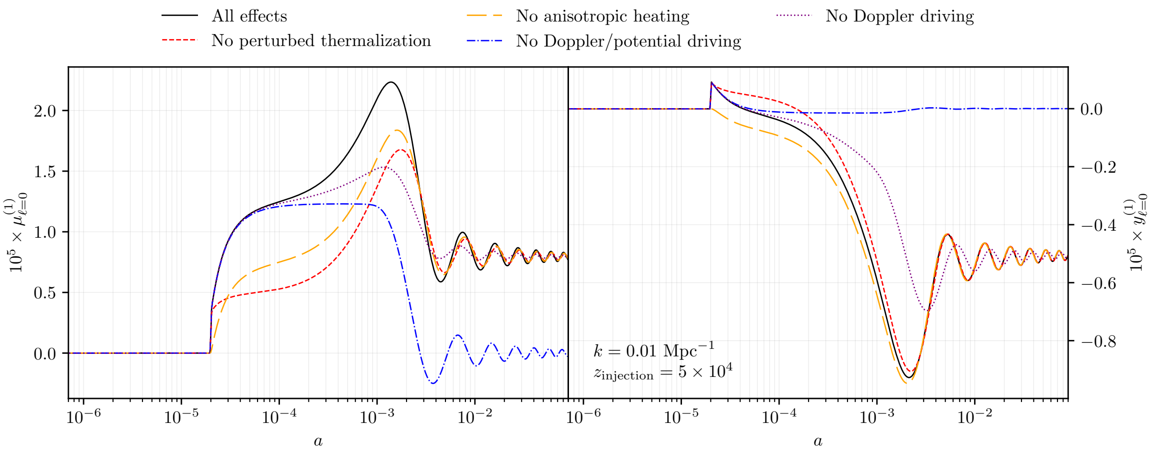


Figure 9: Same as Fig. 8, but focusing on $\mu_0^{(1)}$ (left column) and $y_0^{(1)}$ (right column) for $k = 0.01\,\mathrm{Mpc}^{-1}$. Similar general trends hold to Fig. 6. Perturbed thermalisation and anisotropic heating effects again show equal importance to super-horizon evolution, while boosting is now more important and drives opposite signs for $\mu^{(1)}$ and $y^{(1)}$.

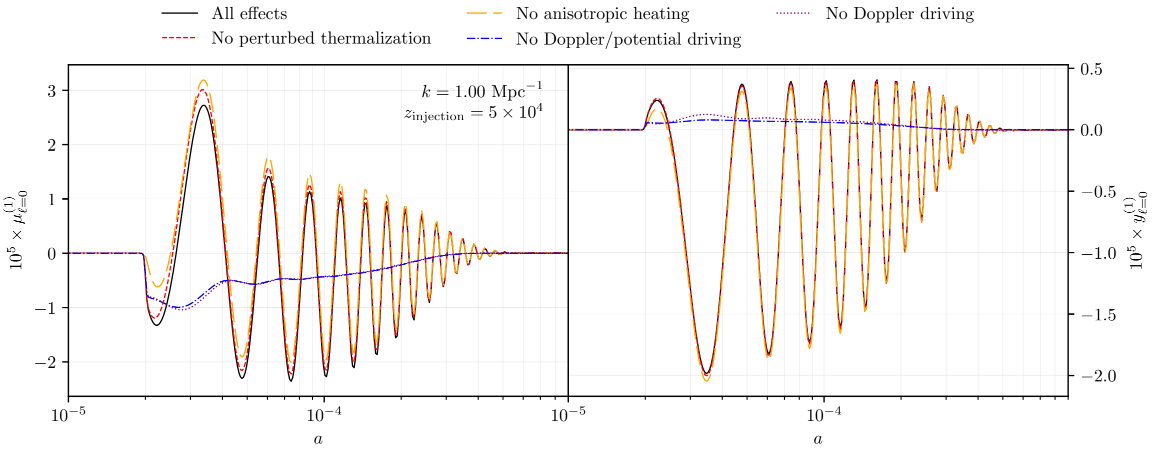


Figure 10: Same as Fig. 9, but for $k = 1$ Mpc^{-1}.

to what we see for average distortions $M_\mathrm{K}\,\tilde{\boldsymbol{y}}_0^{(1)} + \boldsymbol{D}_0^{(1)}$, except with a slowing effect captured in, e.g., the term $\propto \left[\tilde{\delta}_\mathrm{b}^{(1)} + \tilde{\Psi}^{(1)}\right] M_\mathrm{K}\,\boldsymbol{y}_0^{(0)}$. This signifies a small delay to the conversion of y to μ with respect to the average distortion, since $\tilde{\delta}_\mathrm{b}^{(1)} + \tilde{\Psi}^{(1)} < 0$ for adiabatic modes. Because for the considered case, the conversion to μ is extremely rapid, this manifests in a small peak in the μ_0 and y_0 transfer functions before reaching its super-horizon plateau. For later injections, this evolution will be more visible since the conversion from y to μ is less rapid (see Fig. 8).

Focusing on the late evolution, broadly speaking, we can see that aside from minor phase differences the transfer functions of the respective multipoles of all spectral parameters behave similarly. This is expected since the main driver during the late phase is Doppler driving and decaying potentials, which source the distortion anisotropies in very much the same way to the temperature anisotropies. This also means that the distortion-temperature correlations should be significant, as we further demonstrate below.

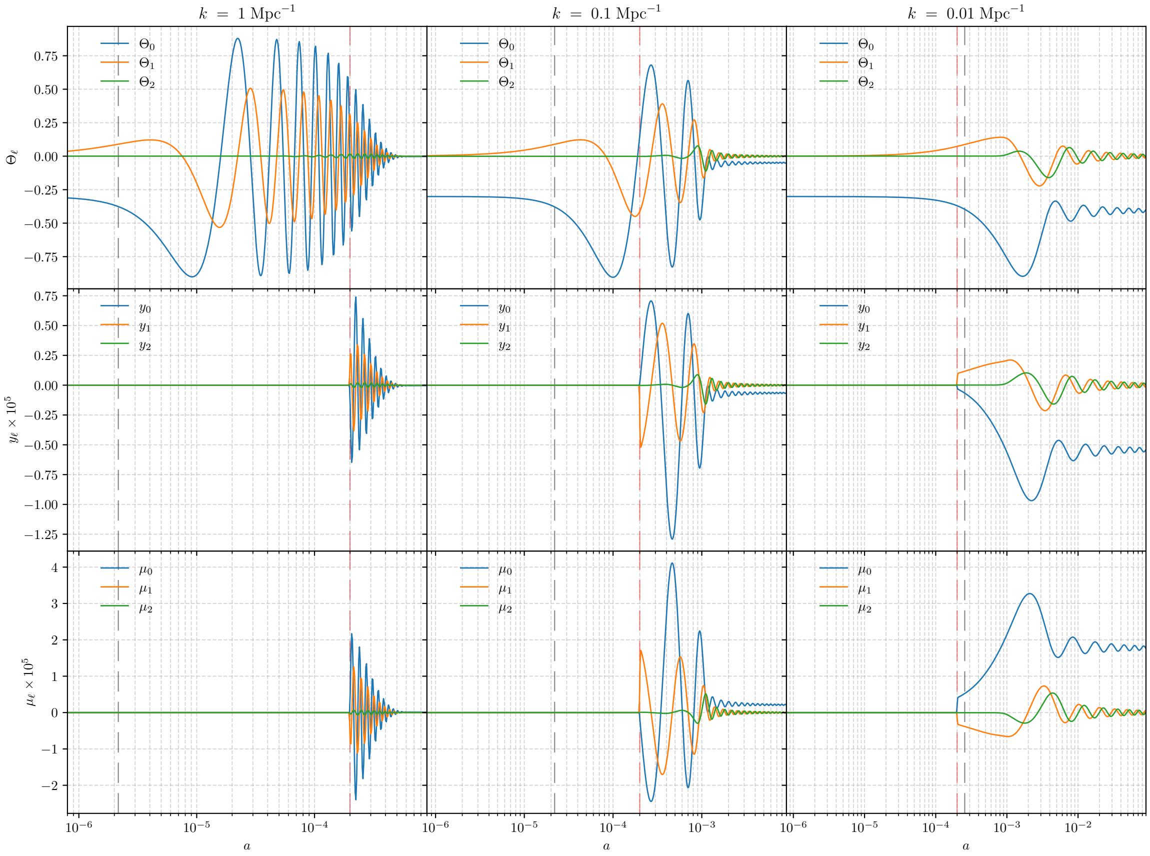


Figure 11: CMB transfer functions for single heating at $z_{\rm injection} = 5{\times}10^3$ with $\Delta\rho_\gamma/\rho_\gamma = 10^{-5}$ and wavenumber k as labeled. This leads to an average distortion that freezes with $\Theta^{(0)} \approx 3.5 \times 10^{-12}$, $y^{(0)} \approx 2.5 \times 10^{-6}$ and $\mu^{(0)} \approx 2.0 \times 10^{-7}$. The numerical solutions are computed using $N_{\rm max} = 15$ with a rotation to the scattering basis (Sect. 2.5). Dashed vertical lines show times of horizon crossing (gray) and energy injection (red).

Regardless of what occurs super-horizon, horizon-crossing will drive a source of both μ and y anisotropies (noticeable shortly after the gray vertical lines). This boosting typically occurs long after the ceasing of thermalisation (for k-modes relevant to CMB power spectra), and will become the dominant sources for late injection (see Fig. 11).

We also mention that one source of y-distortion anisotropies is from the shift in the average CMB temperature by thermalisation. This comes from the Doppler boost of $\Theta^{(0)}$ ($\hat{O}_x G = Y + 3G$) and for the early injection considered here is found to cause $y_0^{(1)} \simeq -10^{-7}\zeta$. At this level, several other terms will become important so that we leave a more detailed investigation to the future. We note, however, that this y-distortion mode could in principle allow us to test changes to the temperature-redshift relation caused at late phases of the cosmic history. To leading order, the expected signal can be thought of as a mismatch of the average CMB spectrum and the spectrum of the CMB anisotropies due to the independent evolution of the average spectrum [37]. In addition, entropy production right after the Big Bang Nucleosynthesis era could be tested, which given current CMB anisotropy constraints on the helium abundance could still accommodate $\Delta\rho_\gamma/\rho_\gamma \simeq 0.01$ [38, 39].

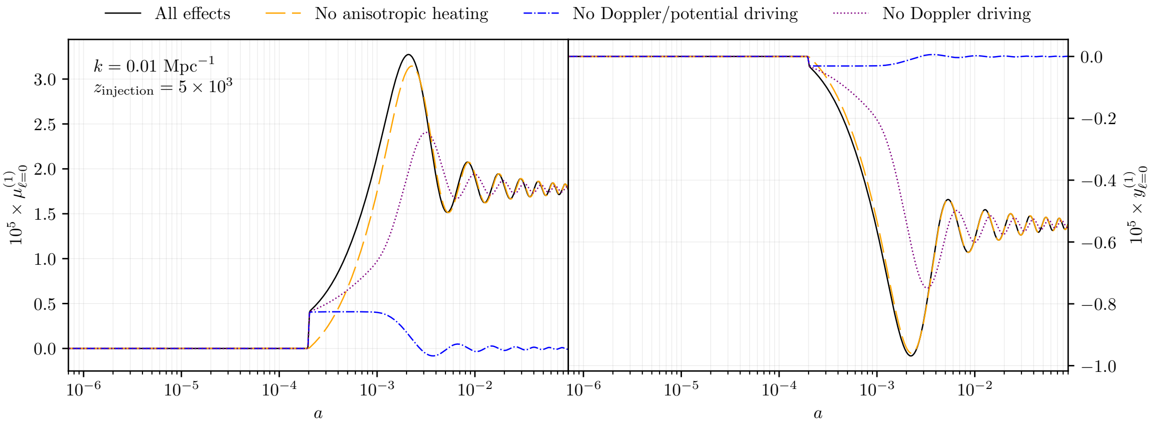


Figure 12: Same as Fig. 11, but focusing on $\mu_0^{(1)}$ (left column) and $y_0^{(1)}$ (right column) for $k = 0.01\,\mathrm{Mpc}^{-1}$. Perturbedd thermalisation is not shown now, since it has no effect (verified independently). Anisotropic heating sources both $\mu^{(1)}$ and $y^{(1)}$ via projections of $Y - Y_1$. Other than this boosting is the dominant source, showing the usual behaviour with the opposite sign mix of $\mu^{(1)}$ and $y^{(1)}$.

3.3 Anisotropies for energy injection in the residual distortion era

We next consider injection at $z_{\mathrm{injection}} = 5 \times 10^4$, an approximate midpoint of the residual distortion era ($10^4 \lesssim z \lesssim 3 \times 10^5$). The average distortion now has a non-vanishing y-distortion contribution – amounting to $\simeq 50\%$ of total energy for this redshift. Needless to say, the transfer functions for $\Theta^{(1)}$ remain unchanged, but are shown again in Fig. 8 for convenience.

The distortion transfer functions all show similar overall behavior as for $z_{\mathrm{injection}} = 5 \times 10^5$ except for some subtle yet notable changes. At $z \lesssim 2 \times 10^5$, DC emission and absorption terms become negligible, like for the average evolution [30]. However, perturbed scattering effects are still relevant, and in comparable to the anisotropic heating (see Fig. 1). The super-horizon evolution now shows both $y^{(1)}$ and $\mu^{(1)}$ contributions from anisotropic heating, however the M_{D} operator together with $(Y_1 - Y)$ causes anisotropic $\mu^{(1)}$ to dominate the picture (see Fig. 9).

Both the μ and y transfer functions become highly correlated around horizon crossing, with boosting now carrying more importance compared to earlier injection times. The mix of both $y^{(0)}$ and $\mu^{(0)}$ at background now produces boosted opposite sign mixes of local $y^{(1)}$ and $\mu^{(1)}$, an effect characteristic of late time injection.

One more small detail we can see in this later injection picture is the effect of injecting a distortion while in sub-horizon evolution, like the case of $k = 1.0\ \mathrm{Mpc}^{-1}$. We can see by comparing the leftmost column of Fig. 8 (compare to Fig. 5) that the oscillations begin immediately following the formation of an average distortion, since many driving sources (e.g. Doppler boosting) are still in effect, and in particular drive with the same frequency in either case. The lack of a super-horizon equilibrium however reduces the noticeable effect of the offset varying mean. Injecting energy close to or after horizon-crossing for smaller k will leave noticeable impacts on the CMB power spectrum, which would be most prominent in the y-era. We see this effect in Sect. 18. To see this clearer we include Fig. 10, where we can explicitly see a lack of contribution from Doppler driving, considering the lack of an average distortion at the time of horizon-crossing. The perturbed thermalisation and anisotropic heating terms are still able to cause a slight offset of the oscillation, but it is much less dramatic than for modes with a full super-horizon phase.

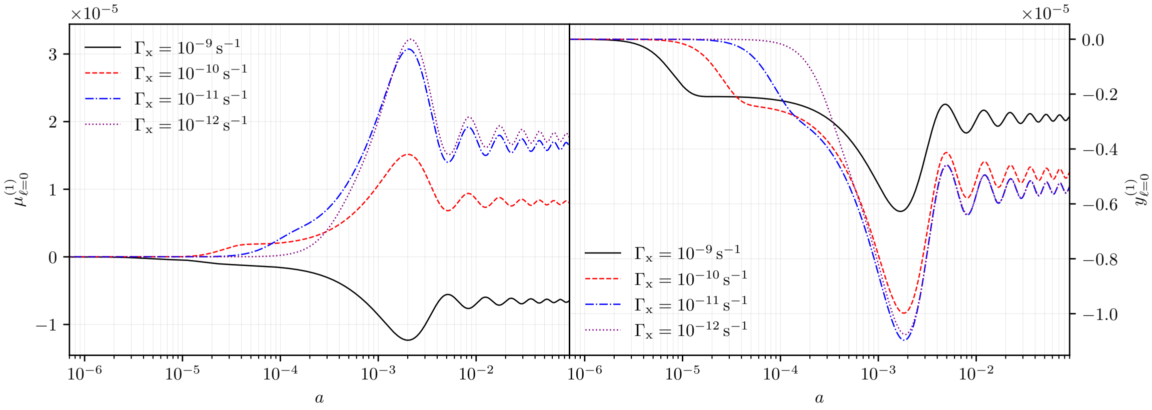


Figure 13: Decaying particle distortion transfer functions for $\mu^{(1)}_{\ell=0}$ (left column) and $y^{(1)}_{\ell=0}$ (right column) at $k = 0.01\,\mathrm{Mpc}^{-1}$ and a total energy release of $\Delta\rho_\gamma/\rho_\gamma = 10^{-5}$. In each figure, we varied the lifetime of the particle. Anisotropic heating is *not* included here. For the short-lifetimes, perturbed thermalisation effects are visible before horizon-crossing.

3.4 Anisotropies for energy injection in the y-era

As a last illustration, we consider distortion anisotropies for injection at $z_{\rm injection} = 5\times10^3$, as shown in Fig. 11. The average distortion is mainly a y-type signal with a $\simeq 2\%$ energy contribution from μ. At this late stage, none of the perturbed thermalisation effects (i.e., scattering and emission corrections) contribute significantly, and the evolution is dominated by the Doppler and potential driving terms upon horizon-crossing. We see by inspecting Fig. 12 that perturbed thermalisation gives an initial boost predominantly μ, but potential driving is the dominant source.

We can see that in all cases shown in Fig. 11 the distortion transfer functions very quickly become highly correlated at a fixed ratio, i.e., $y^{(1)}_\ell \propto \mu^{(1)}_\ell$. This is expected since there is no spectral evolution and the anisotropies simply follow $\hat{O}_x Y$.

The $k = 0.01$ Mpc^{-1} mode crosses horizon very soon after the injection time, and as such still receives the potential boosting contribution. Note however that a smaller k could have undergone gravitational decay before an average distortion existed in the Universe. We will see later that some peaks in the CMB power spectrum are hindered by very late injection time, since they receive Doppler driving but not potential decay (see Fig. 4.3.1).

3.5 Anisotropic heating from decaying particles

All of the discussion presented above only considered an average heating processes at a single redshift. Another interesting case we consider is due to heating by decaying particles, for which two additional aspects become important. Firstly, decaying particle scenarios lead to a more complicated time-dependent evolution of the average distortion [e.g., 25, 32, 40]. This will affect the main distortion transfer functions in interesting ways. Secondly, assuming that the decaying particle densities are modulated by perturbations in the cosmic fluid, anisotropic energy release will occur, which directly creates distortion anisotropies [the $Q'^{(1)}_{\rm c} \in \boldsymbol{Q}'^{(1)}$ term in Eq. (2.1b)]. While the average energy release has been used to constrain decaying particle scenarios based on *COBE/FIRAS* data [26, 32, 41, 42], the latter effect was never before discussed.

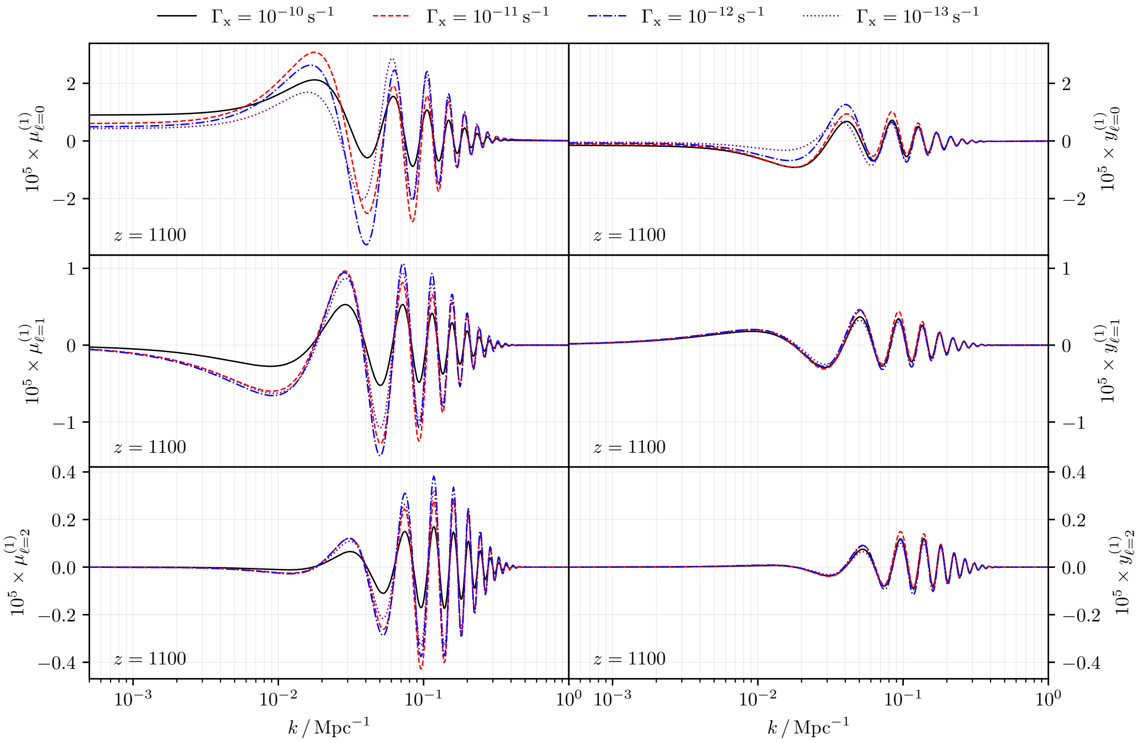


Figure 14: Snapshot of the first three decaying particle distortion transfer functions for $\mu^{(1)}$ (left column) and $y^{(1)}$ (right column) at $z = 1100$ for a total energy release of $\Delta\rho_\gamma/\rho_\gamma = 10^{-5}$. Each row shows a different multipole of the SED amplitude. In each figure, we varied the lifetime of the particle.

3.5.1 Time-dependent heating effect on the distortion transfer functions

Following [25, 43], we implemented a simple heating module for decaying particles, assuming a constant lifetime, $t_X = 1/\Gamma_X$, and mass of the particle, m_X. The average relative heating rate can then be expressed as [see Eq. (6) of 43]

$$\frac{\mathrm{d}\mathcal{Q}^{(0)}}{\mathrm{d}t} \approx \frac{m_X c^2\,\Gamma_X\,N_X}{\rho_\gamma} = \frac{\rho_{X,0}\,\Gamma_X\,\mathrm{e}^{-\Gamma_X t}}{\rho_{\gamma,0}(1+z)} \approx 4.85\times 10^3\, f_{\rm dm}\left[\frac{\Omega_{\rm cdm}h^2}{0.12}\right]\frac{\Gamma_X\,\mathrm{e}^{-\Gamma_X t}}{1+z}, \qquad (3.1)$$

where in the last step we introduced $f_{\rm dm} = \rho_{X,0}/\rho_{\rm cdm,0}$ to allow varying the fraction of dark matter that the particle can make up. Note that calligraphic $\mathcal{Q}$ (as compared to $Q_{\rm c}$) is normalised by $1/\rho_z$, making these expressions match the terms appearing in Eq. (2.1).

In Fig. 13, we illustrate the distortion monopole solutions for various particle lifetimes. We fixed the total energy release to $\Delta\rho_\gamma/\rho_\gamma = 10^{-5}$ by adjusting $f_{\rm dm}$. Comparing to the single injection transfer functions above it is clearly visible how different decay rates smoothly vary across different distortion eras, with shorter (longer) lifetimes having the characteristic final same-sign (opposite-sign) combination of $y^{(1)}_{\ell=0}$ and $\mu^{(1)}_{\ell=0}$ from the boosting effects. This is related to the switch of the early (late) average distortion being M (Y), as discussed in Sect. 2.2. For our illustration we focused on $k = 0.01\,\mathrm{Mpc}^{-1}$, however, the overall picture does not change much when varying k. We also restricted ourselves to decays in the pre-recombination era, such that we could neglect the direct effects of decay on the ionisation history [44, 45]. The latter scenarios can be directly constrained using CMB anisotropies.

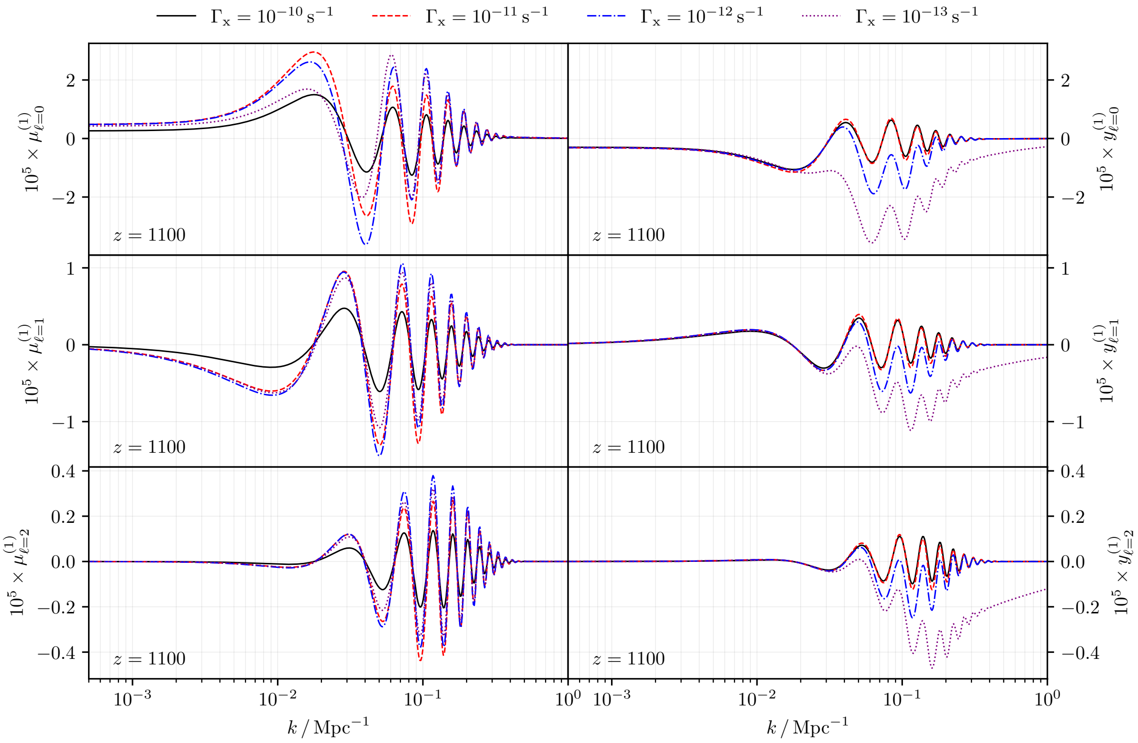


Figure 15: As for Fig. 14 but now perturbed decay was included here, with the effect becoming visible in particular for y_ℓ and late times.

3.5.2 Perturbed decay effect on the distortion transfer functions

In the previous section, we only consider the isotropic part of the heating process. However, if the decaying particle density is assumed to follow the dark matter distribution, we will also have an anisotropic heating term (see Appendix A for a brief derivation).

$$\frac{\mathrm{d}Q_c^{(1)}}{\mathrm{d}t} \approx \delta^{(1)}_{\mathrm{cdm}} \frac{\mathrm{d}Q_c^{(0)}}{\mathrm{d}t}, \tag{3.2}$$

which approximately accounts for the effect of number density modulation that acts alongside the usual modulation of the local time in each Hubble patch $\propto \Psi^{(1)}$ [present for all heating mechanisms as per Eq. (2.1)]. We also assume that heating always only affects the local monopole, sourcing $y^{(1)}_{\ell=0}$.

In Fig. 14, we show snapshots of the monopole, dipole and quadrupole $\mu^{(1)}_\ell$ and $y^{(1)}_\ell$ distortion transfer functions at $z = 1100$, thus highlighting relative contributions to the SD power spectra (see Sect. 4). The lifetime of the decaying particle is varied in each panel. Broadly speaking, the longer the lifetime the larger the contribution of $y^{(1)}$.

In Fig. 15 we show the same figure, with perturbed decay included. For the longest lifetimes we can see a significant enhancement directly from the perturbed decay term (e.g., blue/dashed-dotted lines). This effect is not visible in the $\mu^{(1)}$ transfer function, since at these late times there is no chance for the distortion to thermalise (this could be different for much earlier times than $z = 1100$, however here we are concerned with CMB power spectra).

In contrast to the previous discussions, we find that for the short lifetimes perturbed thermalisation effects contribute noticeably to $\mu^{(1)}_{\ell=0}$ at $k \lesssim 0.01\,\mathrm{Mpc}^{-1}$, and in fact almost cancel the anisotropic heating effects with the perturbed decay included. To illustrate these last two statements more clearly,

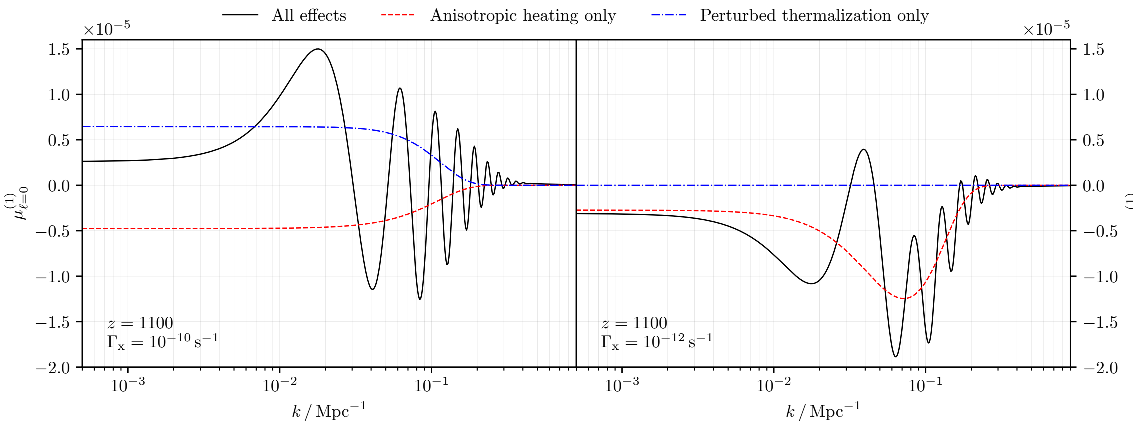


Figure 16: Snapshot of the decaying particle distortion transfer functions for $\mu_0^{(1)}$ (left column) and $y_0^{(1)}$ (right column) at $z = 1100$ and a total energy release of $\Delta\rho_\gamma/\rho_\gamma = 10^{-5}$. The particle lifetime was fixed as annotated but physical contributions were varied. The numerical solutions are computed using $N_{\rm max} = 15$ and the transfer functions are given in the scattering basis.

in Fig. 16 we fixed the lifetimes as annotated but explicitly vary the physics. Perturbed thermalisation decreases rapidly for longer lifetimes (right panel), leaving anisotropic heating as the dominant driving force which enhances the fluctuation amplitude at intermediate and large scales. As the red line in Fig. 16 indicates, this contribution, is quite smooth without acoustic oscillations. In the left panel we see a combined effect of perturbed thermalisation and anisotropic heating almost cancelling, contrary to the intuition built in Fig. 1. This is due to the aforementioned combination of $\delta^{(1)}_{\rm cdm} + \Psi^{(1)}$ which in adiabatic initial conditions evaluate to reverses the sign of the anisotropic heating had we neglected perturbed decay.

We can anticipate that for lifetimes $t_X \gtrsim 10^{12}$ s, the effects may become even more dramatic; however, in this regime also changes to the ionisation history ought to be included. In this case, the Doppler and potential driving effects will reduce, and pure anisotropic heating terms, leading to $y \times y$-type distortions only, will dominate. We will consider this regime in future work.

4 CMB power spectra with primordial distortions

Studying the transfer functions in Sect. 3 has revealed many important physical aspects in the evolution of anisotropic photon spectra: three main types of source connect the average distorted spectrum to local distortions patches. At early times the picture is dominated by anisotropic heating and perturbed thermalisation, with late times seeing main contributions from boosting sources. These local distortion patches undergo their own evolution including Thompson scattering and thermalisation terms yielding complex SED transfer functions.

This all tells us that the simple three-era picture of average spectral distortions does not exist in the anisotropic case, or at least not as directly. Studying the various limiting cases of energy injection into the μ- and y-eras reveals that a mix of both $\mu^{(1)}$ and $y^{(1)}$ will almost always be present in the anisotropic spectrum. It is feasible that by carefully studying the composition of the anisotropic spectrum one could deduce what composition of X, $\hat{O}_x X$, $\hat{\mathcal{D}}_x X$ and $Y_1(x) - Y(x)$ is present, where $X = M(x)$ or $Y(x)$ would give a sense of the origin of the anistropic signal. The bottom line is that there still is a three-era picture, encoded by complex mixes of the SEDs making up the simple picture at background level.

The simplicity lost in the SD description allows us to yield an exciting gain in observational power. Using the formalism described in this paper we can calculate power spectra from the primordial SED perturbations, and thus open the door to apply the conventional tools used in CMB analysis, but now resolving nuanced spectral shapes in place of a simple blackbody. By measuring the cross correlations of temperature with $\mu^{(1)}$ or $y^{(1)}$ we can not only place novel constraints on the total energy release in the primordial plasma, but we can also potentially infer the time of this injection. Furthermore, the precise shape of the spectrum could additionally reveal details of the energy injection itself, with multiple injection or continuous energy release scenarios producing distinct power spectra. We illustrate these points by again studying the range of power spectra arising from single injection events, and contrasting with the particle decay scenarios.

Given the context of observation, the results will now be shown only in the observation basis (see Sect. 2.5). We remind the reader that this projects the $Y_n(x)$ SEDs back to $G(x)$, $Y(x)$ and $M(x)$, making only small use of the residual modes. The observation basis in particular does not preserve photon number in this projection, since the process of finitely binning and observing the frequency space would not allow for number estimates in real observation (see paper I). Because of this, the basis will typically slightly exaggerate the *physical* μ amplitude and compensate with a negative temperature shift, as seen in Fig. 4. However, given the *COBE/FIRAS* constraints on average energy release, the latter is too minor to change the temperature fluctuation significantly, unless more minor (second order) effects would be considered. We are thus left with a marginal boost of μ due to this interplay at the start of the residual distortion era (see [20, 32] and paper I).

Evolving the various SED amplitudes until today can be performed with the usual line-of-sight (LOS) integration by including the modified system in Eq. (2.1). Again we summarise the bottom line from the companion paper II:

$$\tilde{\boldsymbol{y}}_\ell^{(1)}(\eta_f, k) = \int_0^{\eta_f} \mathrm{d}\eta\, g(\eta)\, \tilde{\boldsymbol{S}}_\ell(\eta, \eta_f, k), \tag{4.1}$$

$$\tilde{\boldsymbol{S}}_\ell(\eta, \eta_f, k) = \left[\tilde{\boldsymbol{y}}_0^{(1)} + \tilde{\Psi}^{(1)}\boldsymbol{b}_0^{(0)} + \left(\frac{\partial\tilde{\Psi}^{(1)}}{\partial\eta} - \frac{\partial\tilde{\Phi}^{(1)}}{\partial\eta}\right)\frac{\boldsymbol{b}_0^{(0)}}{\tau'}\right] j_\ell(k\Delta\eta) + \tilde{\beta}^{(1)}\boldsymbol{b}_0^{(0)}\, j_\ell^{(1,0)}(k\Delta\eta) + \frac{\tilde{\boldsymbol{y}}_2^{(1)}}{2}\, j_\ell^{(2,0)}(k\Delta\eta)$$
$$+\left\{\theta_z\Big[M_{\rm K}\,\tilde{\boldsymbol{y}}_0^{(1)} + \tilde{\boldsymbol{D}}_0^{(1)} + \Big[\tilde{\delta}_{\rm b}^{(1)} + \tilde{\Psi}^{(1)}\Big]\Big(M_{\rm K}\,\boldsymbol{y}_0^{(0)} + \boldsymbol{D}_0^{(0)}\Big) + \tilde{\Theta}_0^{(1)}\Big(\boldsymbol{D}_0^{(0)} + M_{\rm D}\,\boldsymbol{y}^{(0)} - \boldsymbol{S}^{(0)}\Big)\Big] + \frac{\boldsymbol{Q}'^{(1)}}{4\tau'}\right\} j_\ell(k\Delta\eta).$$

We note again that the first entry in spectral parameter vector $\boldsymbol{y}_\ell^{(1)}$ is the standard temperature perturbation (see Sect. 4.2). The other SED amplitudes are all smaller in proportion to the total energy injection. Throughout this section we inject total energy $\Delta\rho/\rho = 10^{-5}$, yielding typical dimensionless power spectra of magnitude $\mathcal{D}_\ell^{\Theta\mu} \simeq \mathcal{D}_\ell^{\Theta y} \simeq 10^{-5}\mathcal{D}_\ell^{\Theta\Theta}$. Given $\mathcal{D}_\ell^{\Theta\Theta} \simeq 10^{-9}$ at the largest scales in standard ΛCDM, this implies a typical cross-power spectrum amplitude of $\mathcal{D}_\ell^{\Theta\mu} \simeq \mathcal{D}_\ell^{\Theta y} \simeq 10^{-14}$ in dimensionless units. As we discuss in Sect. 5, this level is in fact just below the sensitivity of *Planck* but already exceeds the sensitivity of *LiteBIRD* and *PICO*.

To compute the signal power spectra one can apply the standard formula

$$C_\ell^{XY}(\eta) = \frac{2}{\pi}\int k^2\,\mathrm{d}k\, P(k)\, \hat{X}_\ell(\eta, k)\, \hat{Y}_\ell(\eta, k), \tag{4.2}$$

where the transfer functions for the variables X and Y are used together with the standard curvature power spectrum, $P(k)$. We shall assume the standard cosmological parameters [2] in all our computations below. We will present results with the usual normalisation $\mathcal{D}_\ell^{XY} = \frac{\ell(\ell+1)}{2\pi}\, C_\ell^{XY}$.

4.1 Numerical setup

The calculation of power spectra using Eq. (4.1) can be numerically challenging. Here we provide details on the new implementation of this calculation within `CosmoTherm`, which relied heavily on the advice provided in section V of [46].

Transfer functions for sufficiently large k undergo Silk damping [5] long before recombination, and do not impact the CMB power spectrum. On the contrary, modes with low k have not yet crossed horizon even at modern times, and thus also have no influence on the CMB spectrum. We therefore limit our calculations to $2 \times 10^{-5} \leq k/\mathrm{Mpc}^{-1} \leq 0.5$, with an understanding that the larger (lower) k in this range impact the high (low) ℓ power spectrum. In particular, we highlight that $k = 0.01\,\mathrm{Mpc}^{-1}$ corresponds roughly to the scale of the first peak in the temperature power spectrum, since it reaches its maximum amplitude at recombination (see e.g. Fig. 4). Most figures in Sect. 3 showed this mode, which can be helpful in observing some of the physical effects discussed below.

One of the largest complicating aspects of the power spectrum calculation is the combination of a slow varying source with a rapidly oscillating bessel function under the same integrand. To illustrate this, we schematically[5] write

$$I = \int_0^{\eta_f} \mathrm{d}\eta\, g(\eta)\, \mathcal{S}(\eta, k)\, j_\ell(k\Delta\eta), \tag{4.3}$$

where we have split the source S from Eq. (4.1) into a source explicitly dependent on perturbed quantities and the leading spherical Bessel function.

Given this decomposition the approach will be to pretabulate a relatively sparse grid of $\mathcal{S}(\eta, k)$ in a relevant region. This greatly simplifies the calculation since the source function varies slowly in log space while also being expensive to calculate – requiring evolving the primordial perturbations forward from much earlier times. The penalty is increased in this new framework where establishing an accurate background spectrum requires solving even background equations from the time of energy injection, long before relevant scales have crossed horizon. The relevant region for this pretabulation is dictated by the visibility function, which in practice can be seen as restricting the integral limits to concentrate around recombination $z \approx 1100$. This is slightly different for the Integrated Sachs-Wolfe effect, whose terms contain an explicit $1/\tau'$ which can be interpreted of as changing $g(\eta) \rightarrow \exp\big(\tau(\eta_f) - \tau(\eta)\big)$ and thus stretching the region of importance all the way to modern times. In the calculations shown below we create a pretabulated region with 500 points $k/\mathrm{Mpc}^{-1} \in [2 \times 10^{-5}, 0.5]$ and 1000 points $\eta/\mathrm{Mpc} \in [200, \eta_0]$, crucially both being log-spaced.

These 2D grids are then interpolated and used for integration with the Bessel functions, which is best done in linear space since the Bessel function zeros are – for our purposes – spaced evenly. To efficiently integrate a highly oscillatory function, in `CosmoTherm` we use Chebyshev integration techniques, and find the integral across η converges with $\simeq 2^{10}$ samples. Another large efficiency boost in the code is to cache the values of the necessary spherical Bessel function[6], knowing that $k\Delta\eta$ falls between 0 and some maximal value $k_{\max}(\eta_0 - \eta_{\min}) \approx 0.5 \times 15000 = 7500$.[7] Finally, it is noteworthy that the power spectrum is a smooth function, and not all j_ℓ need to be integrated. In practice the sampling can become quite sparse towards high ℓ, with a cubic spline making up the missing evaluations.

Following the integral across conformal time we are effectively left with $X(\eta_0, k)$, and the integral across k can be performed as required. Here, the benefit of the pretabulated sources has become

---

[5] Note that to fully express Eq. (4.1) in this form we would take a summation over $j_\ell(k\Delta\eta)$, $j_\ell^{(1,0)}(k\Delta\eta)$ and $j_\ell^{(2,0)}(k\Delta\eta)$ with their corresponding sources, however here our aim is to clarify the computation and will use the simpler expression with a single source.

[6] We use the `Boost`-library `www.boost.org` to accelerate the computation and achieve high precision.

[7] This statement is somewhat cosmology-dependent; however, 7500 is already quite conservative.

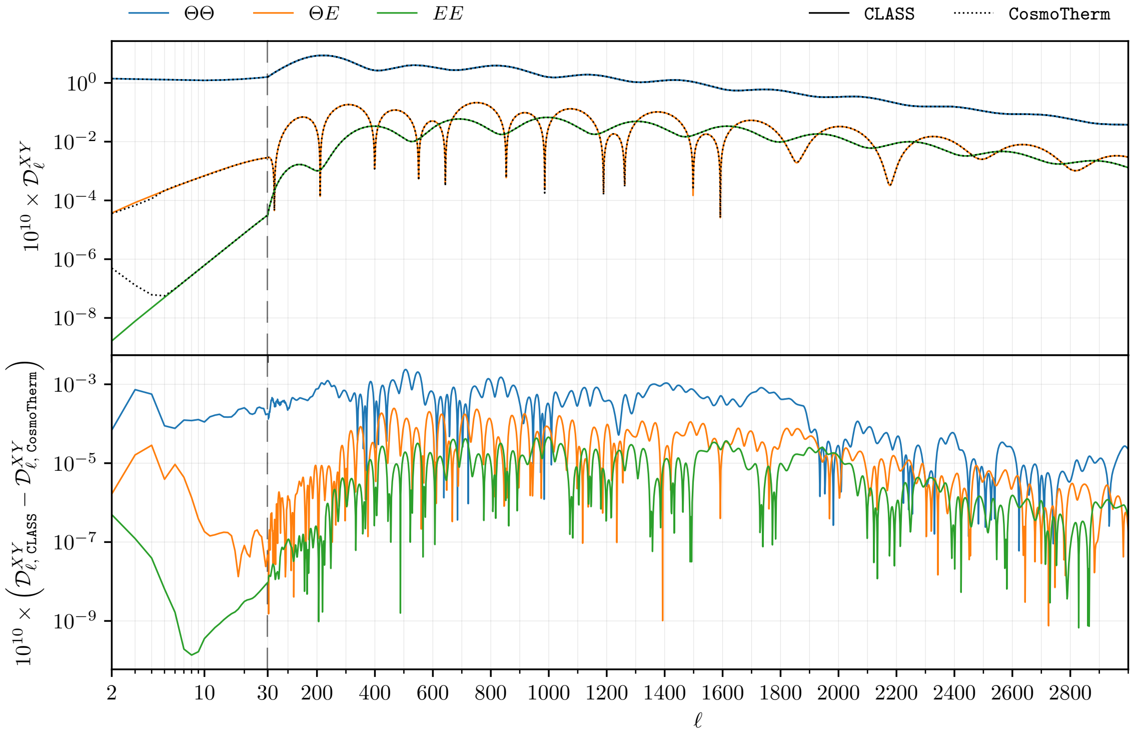


Figure 17: A figure showing the standard, albeit dimensionless, CMB power spectra (ΘΘ, ΘE, EE) both in `CLASS` [28] and `CosmoTherm`. The top panel shows that only qualitative differences exist for very low ℓ in the E mode spectra. The bottom panel reveals through the residuals that differences are below the percent level across the entire ℓ range.

apparent, since many more points are required to effectively capture the oscillations in $X(\eta_0, k)$ than in $\mathcal{S}(\eta, k)$, again because of the spherical Bessel function in the time integrand. Specifically, we calculate $\gtrsim 4000$ points in $X(\eta_0, k)$ from our original grid of only 500 $\mathcal{S}(\eta, k)$ points, and have thus reduced the number of Boltzmann hierarchy calculations by an order of magnitude. This is especially noteworthy in this new treatment of the frequency space, where we have an additional $> (\ell_\mathrm{max} + 1) \times (N_\mathrm{max} + 2)$ equations compared to the standard Boltzmann solvers.[8] For our chosen parameters of $N_\mathrm{max} = 15$ and $\ell_\mathrm{max} = 15$ this amounts to 272 new equations on top of the $5 + 2\ell_\mathrm{max} = 35$ needed for the standard calculation (Φ, δ_cdm, u_cdm, δ_b, u_b, Θ_ℓ, ν_ℓ). Assuming some form of matrix inversion scaling like $\mathcal{O}(N^3)$ we get a solution taking over 1h where it would have previously taken 10s (even an optimistic scaling of $\mathcal{O}(N^2)$ yields a factor of > 10, giving 13m in place of 10s). In this first implementation of the problem, we have had a focus on accuracy and convergence over efficiency, and therefore shall be content with these performance numbers. We find that increasing any parameters here (e.g. ℓ_max and k or η samples) yields no appreciable change to the final results [see however Appendix B for discussion of convergence across N_max]. The efficiency can likely be increased however following more optimization similar to what has gone into state-of-the-art Boltzmann solvers like `CAMB` [27] and `CLASS` [28].

---

[8] In this, the +2 comes from μ and y, while the +1 comes from the fact that the SD sector must be solved at the background level too.

4.2 CMB temperature power spectrum benchmark

The first entry in the photon vector $\boldsymbol{y}^{(1)}$ in this implementation reproduces the `CLASS` $C_\ell^{\Theta\Theta}$ power spectrum to high precision as shown in Fig. 17. The absolute value of relative differences between `CosmoTherm` and `CLASS` amounts to 0.03% for $\mathcal{D}_\ell^{\Theta\Theta}$ and 0.29% for $\mathcal{D}_\ell^{EE}$ once averaging over $2 \leq \ell \leq 3000$ (or 0.02% and 0.03% for averaging residuals without absolute value).[9] These results are achieved in $\simeq$ 30s (wall time) running in parallel over 64 cores, showing some lack of optimisation compared to `CLASS`, however comparable performance can most likely be achieved with further work. We note that the ΘE and EE quadrupole appear much larger in `CosmoTherm` than in `CLASS`, however this will not impact the forecasts considering the cosmic variance at those scales.

4.3 Single injection CMB power spectra

We now have all the ingredients to compute the first CMB parameter power spectra. In Fig. 18, we show the $\Theta \times \mu$, $\Theta \times y$ and $\Theta \times r_1$ power spectra for various injection redshifts and $\Delta\rho_\gamma/\rho_\gamma = 10^{-5}$. A rich acoustic peak structure is revealed, with a clear dependence on the injection epoch.

Starting with late time injection we can see that the peaks in $\Theta\mu$ and Θy are the same shape, with only some negative coefficient relating the two. This is due to boosting as the only source at sufficiently late times, as seen and discussed throughout Sect. 3. This intuition is reinforced by the ratio of observed μ and y energies in Fig. 19, which approaches a consistent value for late time injection. Furthermore the fact that the peaks in the power spectra have similar appearance the usual $\Theta\Theta$ spectrum hints towards the common source of Doppler boosting, which can be verified by inspecting Fig. 20. Finally we note that the low ℓ part of the spectrum is similarly due to the late time ISW effect, again familiar from the standard Cosmological picture.

The earlier times are more complicated, with anisotropic heating and perturbed thermalisation taking on more importance and frequently counteracting the minimal contributions from boosting (see Fig. 6 and compare to top row of Fig. 1). This can be immediately seen by how odd peaks are strongly suppressed in the $\Theta \times \mu$ spectrum, indicating a source which is not governed by Doppler peaks. In fact, the prevalence of even peaks hints towards the effect of baryon loading, variables which partly modulate the local thermalisation efficiency. In the $\Theta \times y$ spectrum the Doppler peaks are still appreciable, a consequence of perturbed scattering favouring the creation of $\mu^{(1)}$ through perturbed thermalisation (through both $M_{\rm D}\boldsymbol{y}_0^{(0)}$ and $Y_1 - Y$) and anisotropic heating thermalising to a $\mu^{(1)}$ spectrum, thus leaving the small boosts as sole contributors to local $y^{(1)}$ distortions.

The amplitude of the $\Theta \times r_1$ power spectrum is roughly one order of magnitude below the $\Theta \times y$, indicating that only about 10% of the SD-energy is contained in this observable. Higher residual distortion power spectra (see Sect. 4.3.2) drop further in amplitude, indicating fast convergence of the signal model and information.

The range of timings varies quite smoothly in the residual-era, but the spectra start to overlap more at the extremes. This implies a strong level of time sensitivity in observation for residual-era injection, while differentiating the moment injection in, say, the y-era will require strong measurements on individual peaks (see discussion in Sect. 5.3). In the case of the μ-era the discriminating power is quite reduced, with peaks mostly overlapping till injection at $z \lesssim 2 \times 10^5$, the moment thermalisation becomes inefficient and the residual-era begins. Nevertheless, a tomographic picture is revealed at $10^3 \lesssim z_{\rm injection} \lesssim 10^5$.

---

[9] We included polarisation effects on the temperature equations but removed reionisation effects from `CLASS` for this comparison.

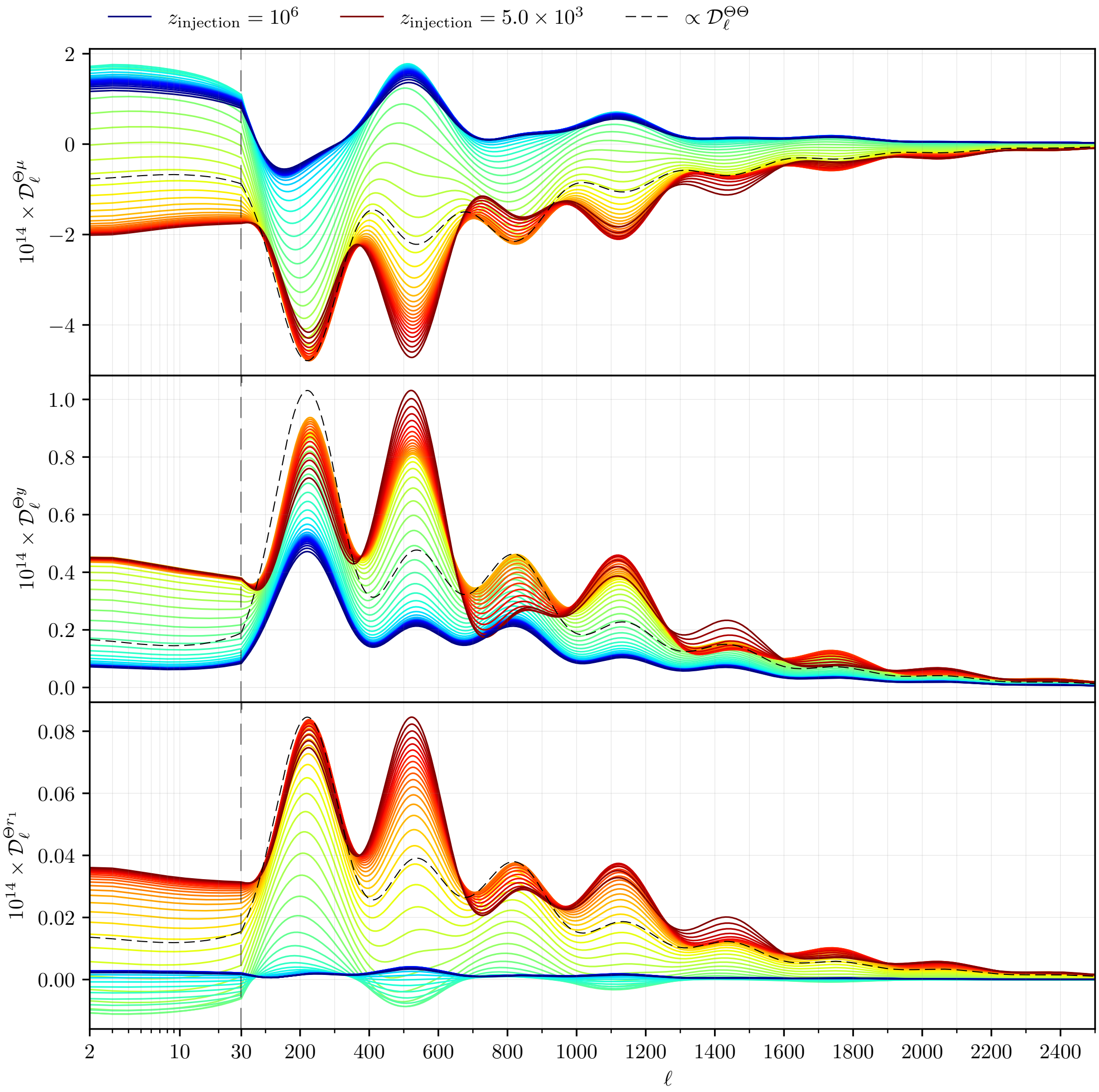


Figure 18: The power spectra for $\Theta \times \mu$, $\Theta \times y$ and $\Theta \times r_1$ over a range of 50 single-injection redshifts. Blue lines show early injection into the μ-era and red lines show late injection in the y-era. The vertical dashed line shows a division between log-spaced ℓ values (left) and linear-spaced values (right). For reference, we show the familiar $\Theta \times \Theta$ power spectrum (rescaled within each panel). Comparing the acoustic peak structure, we recognize that $\Theta \times \Theta$ and the respective $\Theta \times \mu/y/r_1$ power spectra are in phase, a sign of their common origins (e.g., Doppler boosting).

The correlations $y \times E$ and $\mu \times E$ are also important for the forecasts (Sect. 5) and are shown in Appendix D. They are generally more complex and thus less illustrative than the correlations with temperature, hence their omission from main text.

4.3.1 Isolating various physical effects

The power spectra are complex and composite statistics, where each ℓ involves contributions from many k modes which thus encode different times of horizon crossing and thus different relative contributions of aniostropic distortions sources. In order to distil some physics from these data we will

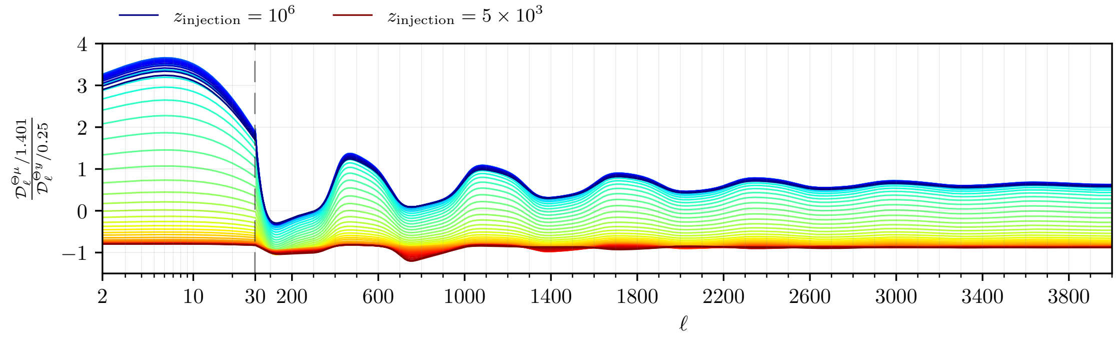


Figure 19: A figure showing ratios of power spectra, illustrating the relative composition of the spectrum in terms of y and μ. The additional factors normalise the amplitudes to their energy content. Both the average change in the μ/y ratio as well as the ℓ-dependent change provide the means to distinguish energy injection scenarios.

again rely on the switches[10] explained in Sect. 3.1.1. We furthermore decompose boosting sources into Doppler boosting from baryon velocities and gravitational potential decay (the latter contributing mostly to late time ISW effects).

These switches allow us to dissect the rich features in the acoustic peaks themselves. To isolate a physical effect we calculate the power spectrum with and without the relevant terms in the evolution equations, and plot the difference between the two. For example, the Doppler contribution is found by subtracting the solution without Doppler driving from the full solution.[11]

The main point Fig. 20 illustrates is that Doppler driving is the dominant effect on the SD power spectra, with only early injection times seeing another comparable term. At these early times we have already seen that anisotropic heating and perturbed thermalisation become large contributors to the SD signal. Potential driving terms are most important at large scales ($\ell \leq$ 30–40), introducing an integrated Sachs-Wolfe plateau to distortion signals. Although less important at small scales (high ℓ), the potential driving terms provide important time-dependent information.

One notable feature in the single injection scenario is that for the latest of injection times the first peak starts to wane while the second peak continues its growth. The turn over point in the first peak happens around $z_{\rm injection} \approx 2 \times 10^4$. Similarly we see the third and fourth peak affected by the late injection. We can see that these changes are primarily caused by changes in the potential driving late into the y-era. Starting with $z_{\rm injection} = 5 \times 10^3$ we see that potentials don't drive $\ell > 1000$, which received contributions from k-modes which were deep into the horizon at the time of injection, and thus saw almost no potential driving. Even for $\ell < 1000$ we see smaller potential effects with decreasing injection redshift, which are likely caused by some combination of the aforementioned effect *spreading* over ℓ and the fact that potential decay is greatly reduced close-to and beyond the matter-radiation transition. These potential decay effects are also visible, although less clearly, in the transient effects on the monopole and dipole transfer functions in the central column of Fig. 8 (injection near horizon crossing) and Fig. 11 (sub-horizon injection). This is in contrast to Fig. 5 where the same mode received large boosting from potential decay at the time of horizon crossing, since the average SED amplitudes had been sourced prior.

---

[10] We chose to leave the temperature equations unchanged and *only* switch distortion drivers

[11] There is no way of showing the true isolated effects since the power spectrum is a squared statistic, and thus no simple superposition principle can be used. This technique however is highly illustrative.

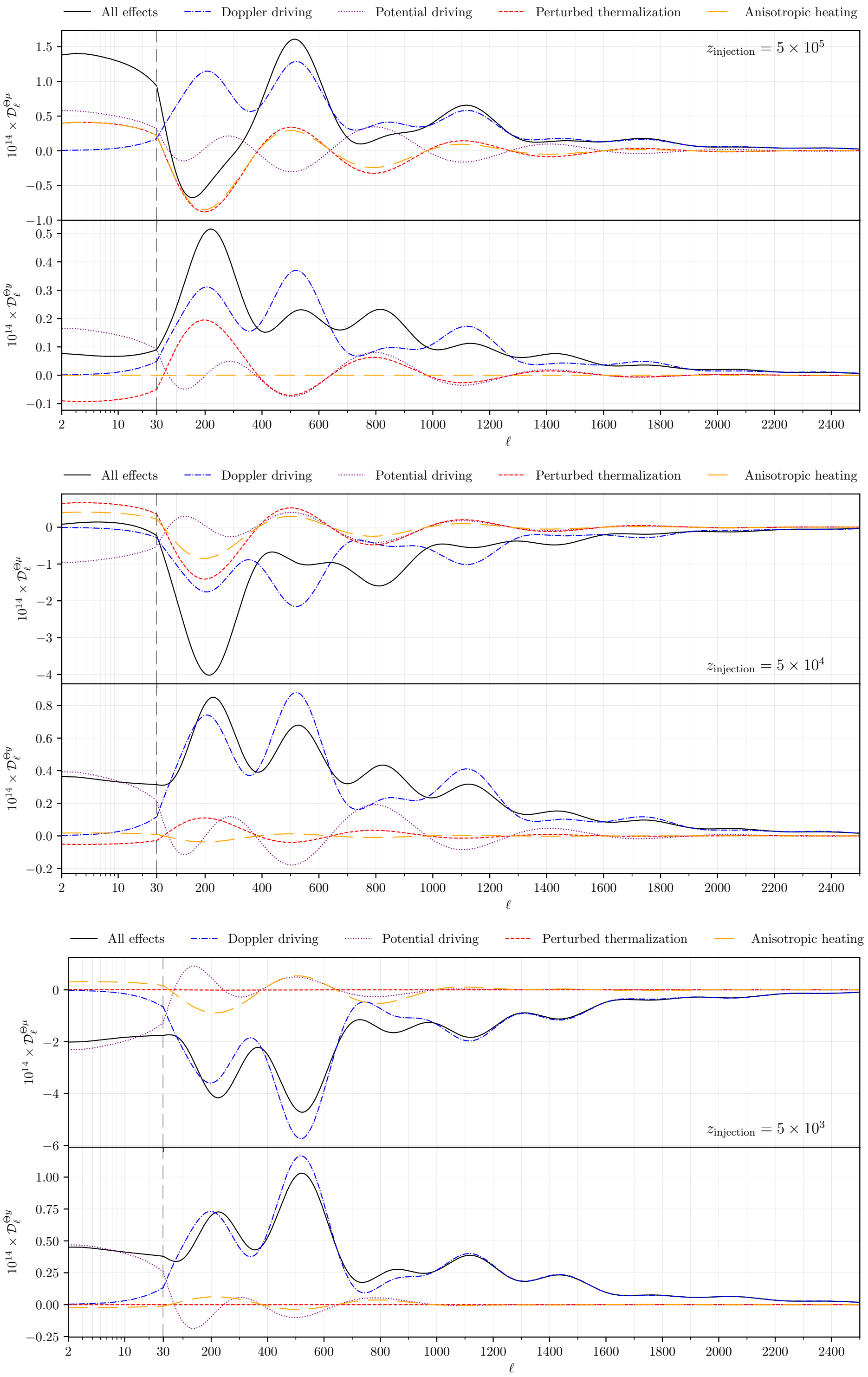


Figure 20: Three figures showing the power spectra for $\Theta \times \mu$ and $\Theta \times y$ for the three usual single-injection redshifts corresponding to each main SD era. Different lines indicate the inclusion or exclusion of a physical term (made more explicit in main text). A vertical dashed line shows a division between log-spaced ℓ values (left) and linearly-spaced values (right).

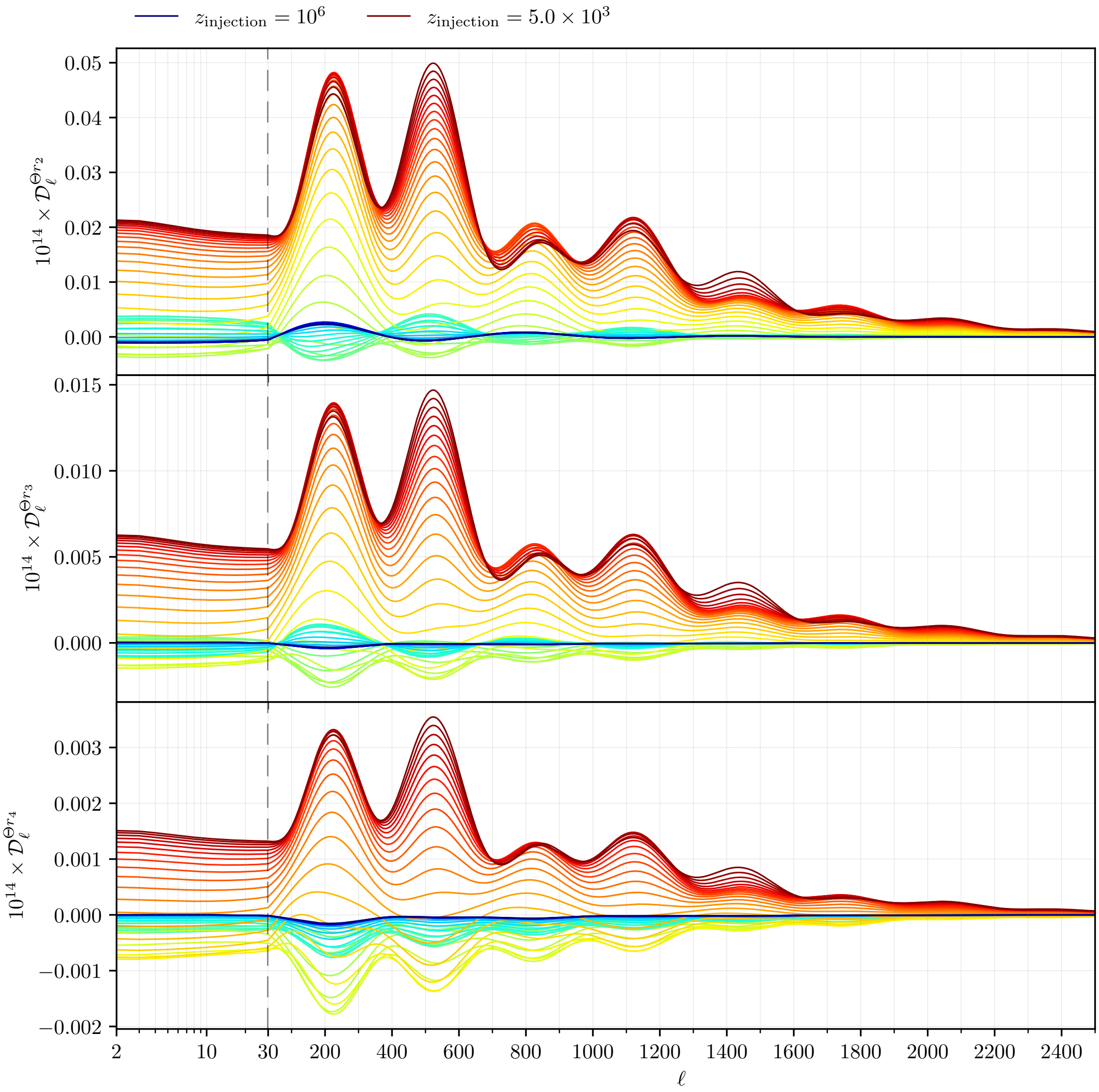


Figure 21: As for Fig. 18 but this time showing correlations between Θ the higher residual distortion modes, r_2, r_3 and r_4.

The anisotropic heating contributions enhance μ at early injection times, and a mix of both μ and y for all other times. While this follows the conventional picture in the residual-era – energy thermalises to some intermediate spectral shape – it is initially surprising for the late injection times. This is due to the additional anisotropic heating term we identify within $\boldsymbol{S}^{(0)}$ (see Sect. 2.2), which sources a spectral shape corresponding to $Y_1(x) - Y(x)$, thus having a nonzero projection onto $M(x)$.

By individually switching perturbed emission and perturbed scattering (not shown) we can confirm that emission is only ever a small subdominant contribution for the injection times considered here, and furthermore the dominant part of perturbed scattering is the $M_{\rm D}\boldsymbol{y}^{(0)}$ and $(Y_1 - Y)$ terms, with the other terms simply providing a delaying effect on the natural thermalisation local anisotropies undergo.

4.3.2 Higher-residual power spectra

In Fig. 21 we present the cross correlations for $\Theta \times r_2/r_3/r_4$, which show the amount of information not captured by the simple decomposition shown above. Importantly the residual modes are rank ordered by their relative importance as can be seen by the decreasing amplitude (they are all normalised similarly to the y distortion, with a relative energy density $E_{r_n} = r_n/4$).

Interestingly, we see how $\Theta \times r_2$ and $\Theta \times r_3$ follow a similar growing shape to the $\Theta \times y$ spectrum for late times. This can be understood by considering that the dominant signal source of power spectra is often the Doppler driving term (see Fig. 20), and upon studying boosts of $Y(x)$ around two residual modes are required for a good fit (see Fig. 2). In a similar way the μ-era injection makes less use of residual modes, a fact which relates to the decreased importance of the boosting sources. We see the residual modes amplitude drop by around a factor of 5 for increasing r_n, showing the decreasing contributions. The remaining energy content in r_4 is quite small, a fact which relates also to convergence within the basis – we notice small, albeit non-negligible changes in the amplitude of r_4 when increasing from, e.g., $N_{\rm max} = 13$ to $N_{\rm max} = 15$, which currently is close to the limits of our computation. This all hints towards the statement that using roughly 6 numbers (Θ, μ, y, $r_{1/2/3}$) is enough information to fully parameterise the photon spectra in the basis chosen here, however $\simeq 18$ are needed in the computation basis to capture the evolution in the most difficult regimes. This statement is, of course, basis dependent (see Sect. 2.5), and specifically it does not exclude the possibility of finding an optimised smaller basis for given energy release scenarios and eras. We discuss this possibility and implications in Sect. 5.2.

4.3.3 Distortion auto-power spectra

Although they are far below the detection prospects of even future imagers (see Sect. 5) it is illustrative to study the purely SD power spectra. In Fig. 22 we show the $\mu \times \mu$, $y \times y$ and $\mu \times y$ spectra, together with a rescaled $\Theta \times \Theta$ spectrum for comparison.

The auto-power spectra show almost exactly the same structure for late injection times (upto an overall scale) since dominant source is boosting, yielding a fixed ratio of of $y^{(1)}$ and $\mu^{(1)}$ amplitudes regardless of the ℓ. An extension of this is that that the cross spectrum $\mu \times y$ shows a similar shape but with a negative sign, since the boost of $Y(x)$ matches opposite sign mixes of the $y^{(1)}$ and $\mu^{(1)}$.

The first and third peak in $\Theta \times \Theta$ spectrum appears to have no corresponding peak in the SD case. The effect is actually slightly exaggerated – if the early injection times were amplified for $\Theta \times y$ (blue line in the middle row) then the peaks would in fact be present with the expected ratios, but not for late time injection. In Figs. 34, 35 and 36 in Appendix C we show the effects of physical switches on the distortion spectra in the three characteristic eras. Those figures suggest that this loss of peaks for late time injection occurs due to the missing potential driving terms from energy injection close to horizon-crossing.

Interestingly while there is a strong correlation of $\Theta \times y$ for early injection times, there is a very low correlation of $y \times y$ distortions and a complex pattern of $\mu \times \mu$. In particular at the lowest ℓ we see greatly enhanced $\mu \times \mu$ since the super-horizon sources favour production of $\mu^{(1)}$ distortions. The first feature in the $y \times y$ is associated with boosting of those modes at horizon crossing, before which there are no strong sources of anisotropic $y^{(1)}$ [see overall scales in Fig. 6].

4.4 Decaying particle CMB power spectra

In Fig. 23 and Fig. 24, we show the CMB power spectrum for various particle decay lifetimes, where in the latter figure we include effects of perturbed decay (see Sect. 3.5). Without perturbed decay the curves resemble those seen in Sect. 4.3, showing that the single injection scenario, while unphysical,

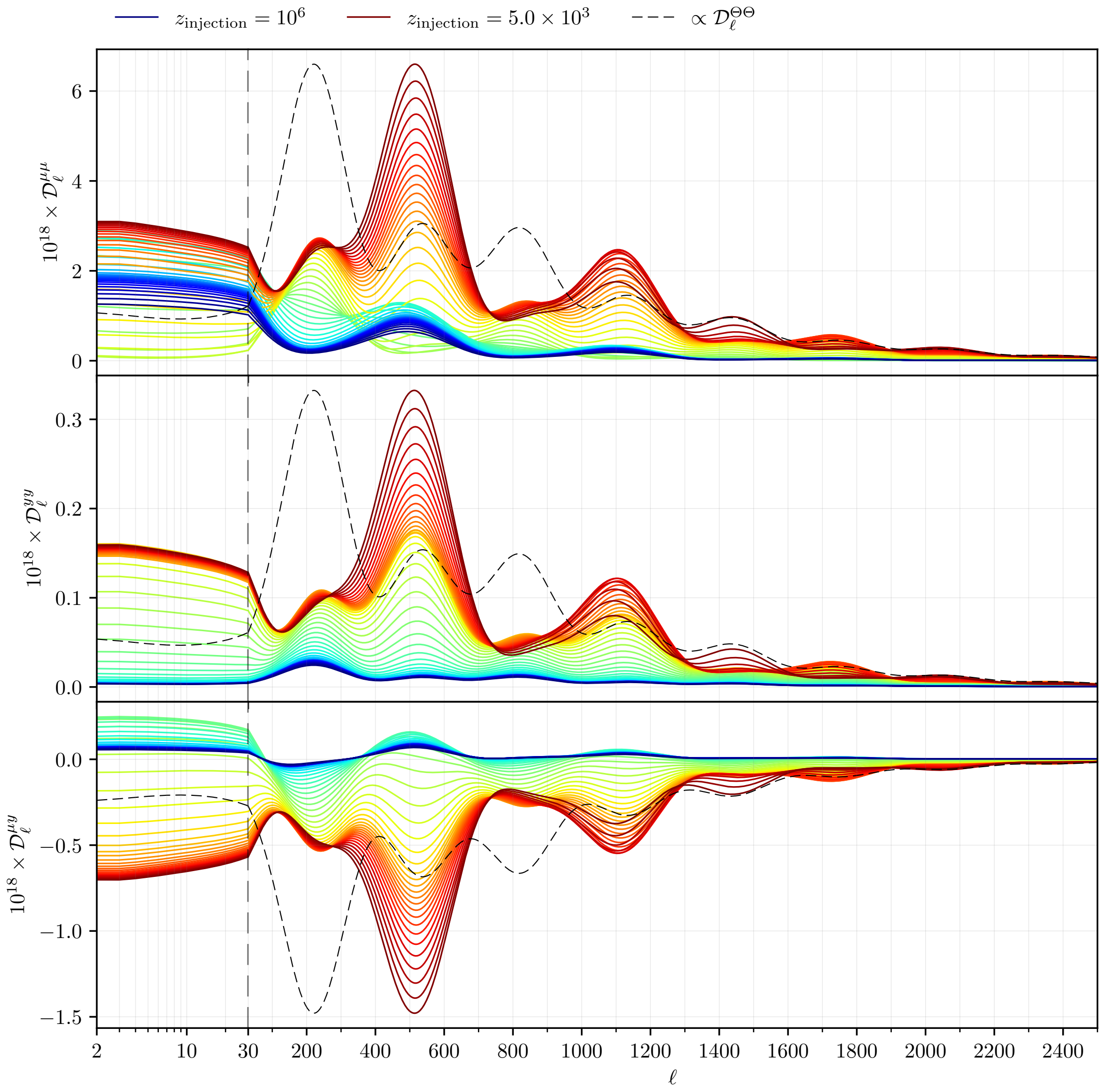


Figure 22: A similar figure to Fig. 18 but now showing the $\mu \times \mu$ (top), $y \times y$ (middle) and $\mu \times y$ (bottom) spectra. Shown for comparison is the (re-scaled) dimensionless temperature power spectrum.

serves as a good illustration of realistic continuous energy injection scenarios if the window of energy creation is sufficiently narrow. The perturbed decay on the other hands changes both the early and late injection scenarios. Due to the adiabatic initial conditions, injection in the μ-era sees a partial cancellation between the new $\propto \delta_{\rm cdm}$ term and the $\propto \Psi$ term within the usual anisotropic heating. This allows boosting to take a more central role in the formation of the anisotropic spectrum, and thus a more dominant first peak in the power spectrum (see Fig. 20). However, the biggest notable feature is the enhancement (reduction) of the odd (even) peaks in the late injection time $\Theta \times y$ spectrum.

The effect of perturbed decay on the spectra is well illustrated by again taking a ratio of the relative y and μ energy densities as seen through their cross correlation with temperature fluctuations. This is shown in Fig. 25, where the bottom panel indicates a large enhancement towards the y energy density beyond $\ell = 1000$. This is understood since the perturbed decay injects energy directly into y,

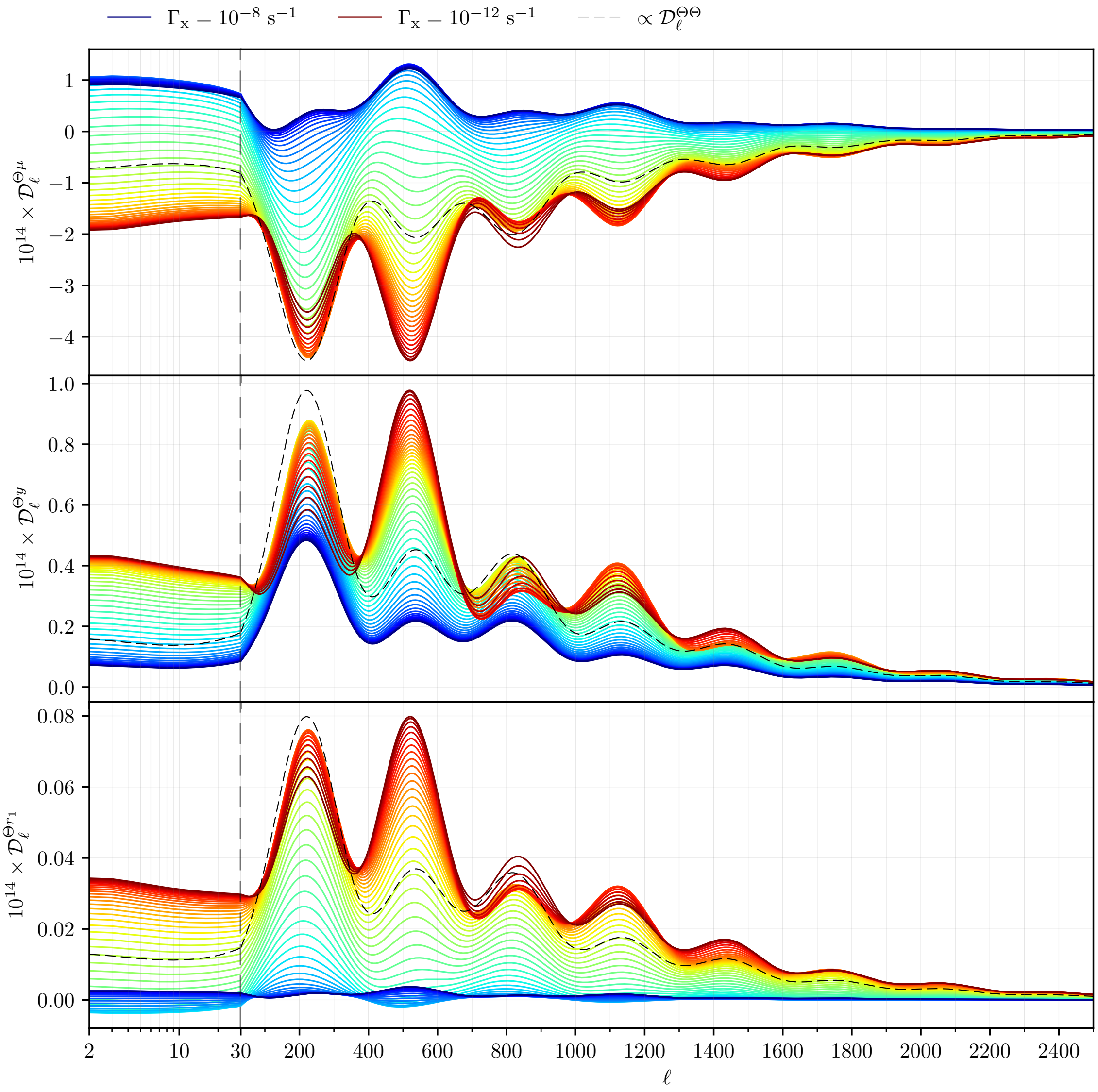


Figure 23: A figure showing the power spectra for $\Theta \times \mu$ (top panel) and $\Theta \times y$ (bottom panel) over a range of decaying particle lifetimes. Blue lines show short lifetimes, thus decaying predominantly in the μ-era, while red lines show long lifetimes, therefore decaying predominantly in the y-era. A vertical dashed line shows a division between log spaced ℓ values (left) and linearly spaced values (right).

which has no time to boost into a mixed spectrum for the later injection scenarios. This model serves as a motivating example and an enticing hint that a powerful future probe of concrete energy injection mechanisms could be to detect specific enhanced peaks in CMB power spectra.

5 Fisher forecasts

To assess the detectability of the signal and have a mean to compare the prospective constraints on energy injection to the *COBE/FIRAS* [47, 48] limits we use a Fisher matrix forecast, which allows us to quickly set a lower bound on parameter errors for a given instrumental configuration. Here we consider a simplified scenario where the only free parameter is the fractional injected energy

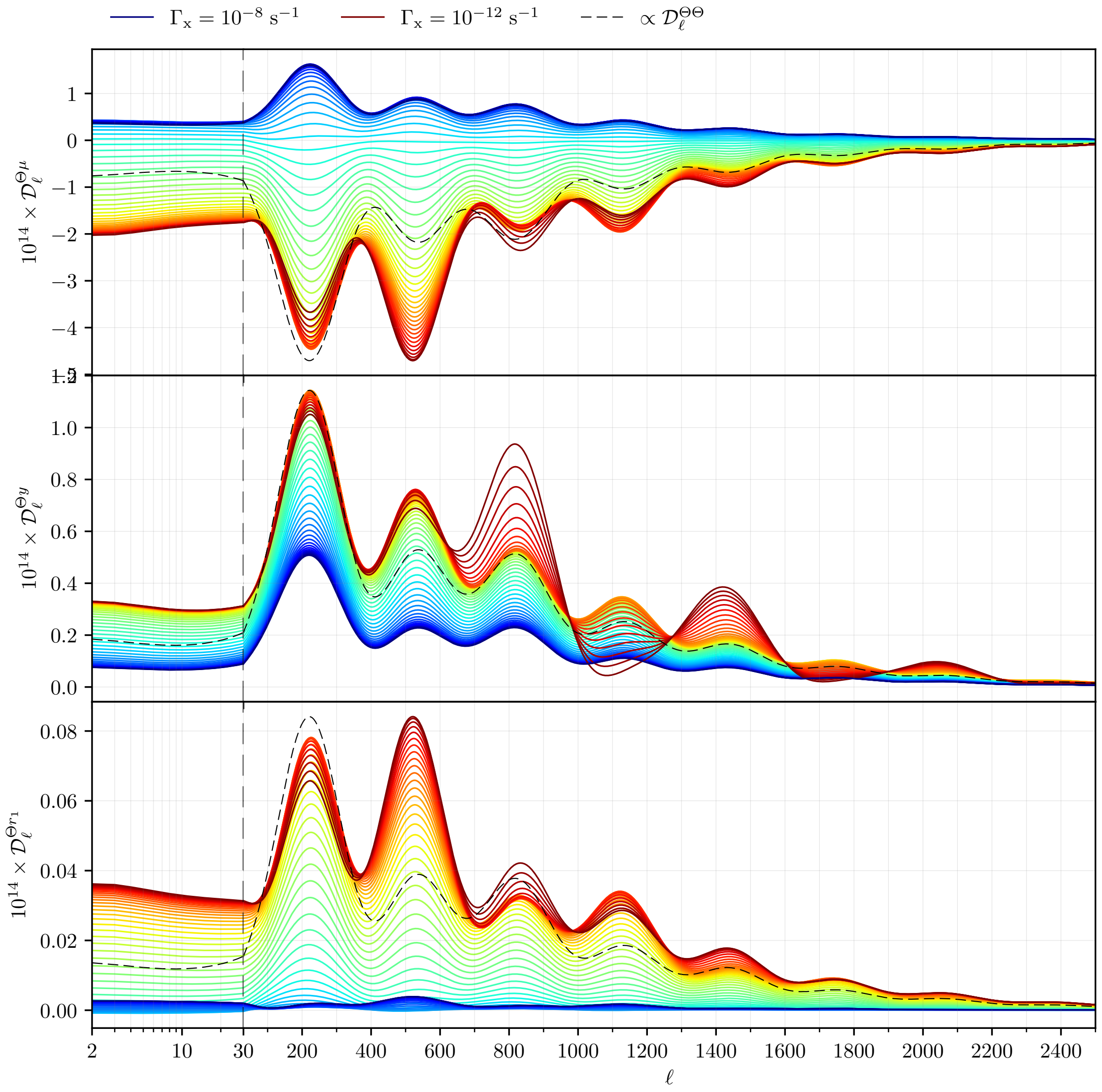


Figure 24: Same as Fig. 23, but now including an anisotropic heating term from the perturbed decay of particles modulated by local matter densities.

$\Delta\rho/\rho$, while all other cosmological parameters and remaining energy-release-model parameters (e.g. redshift of injection or decaying particle lifetime) are fixed.

As observables we consider using all the cross correlations between spectral distortions μ and y and CMB primary anisotropies T and E, neglecting the residual distortion contributions. In this case, the estimate of the $\Delta\rho/\rho$ error reads

$$\sigma_{\Delta\rho/\rho} = \left[\sum_\ell \left(\partial_{\Delta\rho/\rho}\hat{C}_\ell\right)^T \Sigma_\ell^{-1}\,\partial_{\Delta\rho/\rho}\hat{C}_\ell\right]^{-1/2}. \tag{5.1}$$

Here $\hat{C}_\ell = \left(\hat{C}_\ell^{\mu T}, \hat{C}_\ell^{\mu E}, \hat{C}_\ell^{yT}, \hat{C}_\ell^{yE}\right)^T$ is a vector of the observable spectra. To build our intuition we will also show partial results that involve only a subset of spectra; those cases are produced by simply

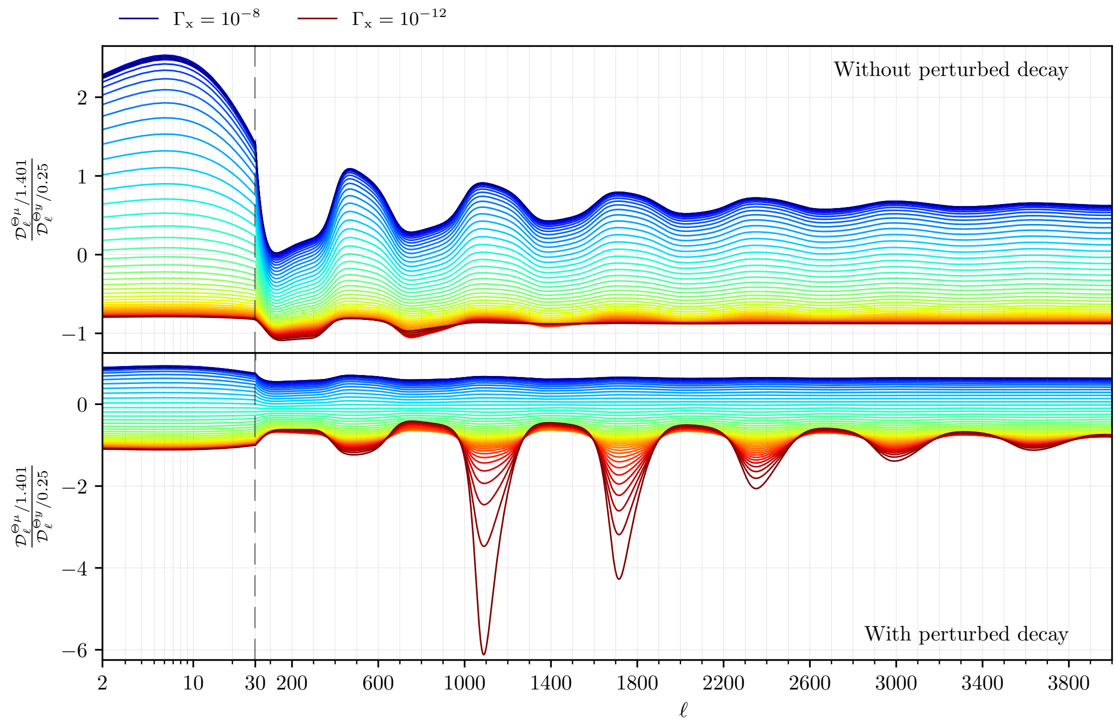


Figure 25: As for Fig. 19, but now for the case of particle decay. Importantly the bottom panel shows the effects of perturbed decay, which affects the peaks $\ell > 1000$.

removing the irrelevant entries from $\hat{C}_\ell$ and from their covariance matrix Σ_ℓ.[12]

In principle, additional information on the fractional injected energy could be extracted from the spectral distortion auto and cross-correlations. However, in a real world scenario they are too faint compared to noise and foregrounds to be measured successfully.

To compute the errors, we use the power spectra from the previous sections. Those were all computed using $\Delta\rho/\rho = 10^{-5}$, but since the cross-power spectra considered here simply scale linearly with $\Delta\rho/\rho$, the derivatives in Eq. (5.1) are trivially obtained. We specify that all limits shown here are calculated assuming a non-detection of the spectra in question. The elements of the covariance matrix have the form

$$\Sigma\left(\hat{C}_\ell^{\alpha a}, \hat{C}_\ell^{\beta b}\right) = \frac{1}{f_{\rm sky}(2\ell+1)}\left(\hat{C}_\ell^{\alpha\beta}\hat{C}_\ell^{ab} + \hat{C}_\ell^{\alpha b}\hat{C}_\ell^{\beta a}\right). \tag{5.2}$$

We model each component as $\hat{C}_\ell^{\alpha\beta} = C_\ell^{\alpha\beta} + N_\ell^{\alpha\beta}$, where the first terms are the theoretical spectrum previously calculated and the N_ℓ are the Constrained Internal Linear Combination (CILC) [49] noise that we will now discuss.

To simulate the impact of foregrounds and instrumental noise on the cross correlations recovered from actual maps, we employ the method outlined in [50, 51] according to the implementation of [52], to which we refer for the details. Working at power spectrum level, we write, for any ℓ, the inter-frequency-channel covariance as sum over instrumental noise, foregrounds and cosmological signals $C_{\ell,\nu\nu'} = N_\ell^\nu \delta_{\nu\nu'} + \sum_{i\in\text{foregrounds}} C^i_{\ell,\nu\nu'} + \sum_{i\in\text{signals}} C^i_{\ell,\nu\nu'}$. The Kronecker-$\delta$ encodes the fact that we take the instrumental noise to be uncorrelated across different channels; the foregrounds encompass dust, synchrotron, free-free, radio and infrared sources [51, 53–55]; the signal are again the ones described previously and we model them as perfectly correlated at all frequencies. To relate the noise and foreground SEDs to the adimensional theoretical spectra we convert them in thermodynamic units with the standard relation $\Theta = c^2 \mathcal{G}(x)^{-1} \nu^{-3}/(2h)\, \Delta I_\nu$, using $T_{\rm CMB} = 2.7255$ K in the conversion. For the SD contributions, this transformation does not remove the frequency-dependence, which is accounted for in the component separation process [e.g., 56]. As it is now well known [57], deprojecting different spectral shapes is essential to obtain unbiased spectral measurements. Following the rationale of deprojecting stronger signals from the fainter maps, we consider the noise contribution to the temperature power spectrum as obtained with the standard ILC

---

[12] We point out that here we implicitly disregarded ℓ couplings even thought they would be non-negligible in an actual survey due to masking and foregrounds. This will be discussed in detail with the analysis of the *Planck* maps in future work.

$$N_\ell^{TT} = \left[\mathcal{G}(\nu)\, C^{-1}_{\ell,\nu\nu'} \mathcal{G}(\nu')\right]^{-1}, \tag{5.3}$$

the y (tSZ) spectrum as obtained with CILC deprojecting T, and μ deprojecting both T and y, i.e.

$$N_\ell^{yy} = \left[(\mathcal{Y}_0, \mathcal{G})(\nu)\, C^{-1}_{\ell,\nu\nu'}\, (\mathcal{Y}_0, \mathcal{G})^T(\nu')\right]^{-1}_{0,0}, \quad N_\ell^{\mu\mu} = \left[(\mathcal{M}, \mathcal{Y}_0, \mathcal{G})(\nu)\, C^{-1}_{\ell,\nu\nu'}\, (\mathcal{M}, \mathcal{Y}_0, \mathcal{G})^T(\nu')\right]^{-1}_{0,0}. \tag{5.4}$$

While the cross correlations like $N_\ell^{\mu y}$ might be important if we were considering the related spectrum as an observable to be analyzed, for which they could constitute a bias [58], they are subdominant in the covariance and thus neglected.[13] Likewise, we consider the temperature and polarisation power spectra to be de facto cosmic variance limited. We apply the methodology just described to *Planck* [59], which represent the current state of the art, *LiteBIRD* [60] as a near-future advancement, and *PICO* [61] as more futuristic scenario. We summarize the instrument specifications we use in appendix E. In all cases we conservatively assume $f_{\rm sky} = 0.65$, and set the maximum ℓ in the sum in Eq. (5.1) high enough to saturate the constraints.

The surveys considered here rely on differential measurements, therefore it is worth noticing that an accurate inter-channel calibration is needed to prevent biasing the measured spectra [57, 62]. However, in each case, the designed accuracy is low enough to prevent a bias significant compared to the survey sensitivity [57]. A precise modelling of the frequency channels band-pass was found to not be fundamental in [58]. Therefore, neither of these two effects will be included at this stage. Future studies should also consider other frequency dependent anisotropic signals, like the Rayleigh scattering at recombination, or resonant scattering with metallic elements in the dark ages [see discussion in Sect. 2.7 of 63]. These other signals constitute additional foregrounds that need to be separated due to their distinct distortion shape. In the case of the Rayleigh scattering, which is the brightest of the aforementioned signals [64], the peculiar ν^4 spectral shape would be easily distinguishable during component separation. Moreover, such spectral dependence also means that at low frequency, where most of the signal is concentrated for both μ and y, Rayleigh scattering is negligible. Therefore we choose to not consider the Rayleigh scattering foreground here. The details of such a component separation however depends on the quantity and distribution of frequency channels, and warrants future work.

In Fig. 26 we show (blue) the $\Theta \times \mu$ cross-correlation for three different injection times and a total energy release of $\Delta\rho/\rho = 3 \times 10^{-5}$, compatible with the 1σ *COBE/FIRAS* limit. That has to be compared with (green) the square root of the covariance element as defined in Eq. (5.2). For reference we compare the signal to (orange) the "standard" calculation for $\Theta \times \mu$ from primordial non-Gaussianity [65, 66] with $f_{\rm NL}^{\rm loc} = 3000$, close to the 1σ *Planck* limit [58]. We can appreciate that with these specific values of $f_{\rm NL}^{\rm loc}$ and $\Delta\rho/\rho$ they are comparable in amplitude.

---

[13] In fact in [58] it was found that using de-projected maps $N_\ell^{\mu T}$ is negligible, even as a bias.

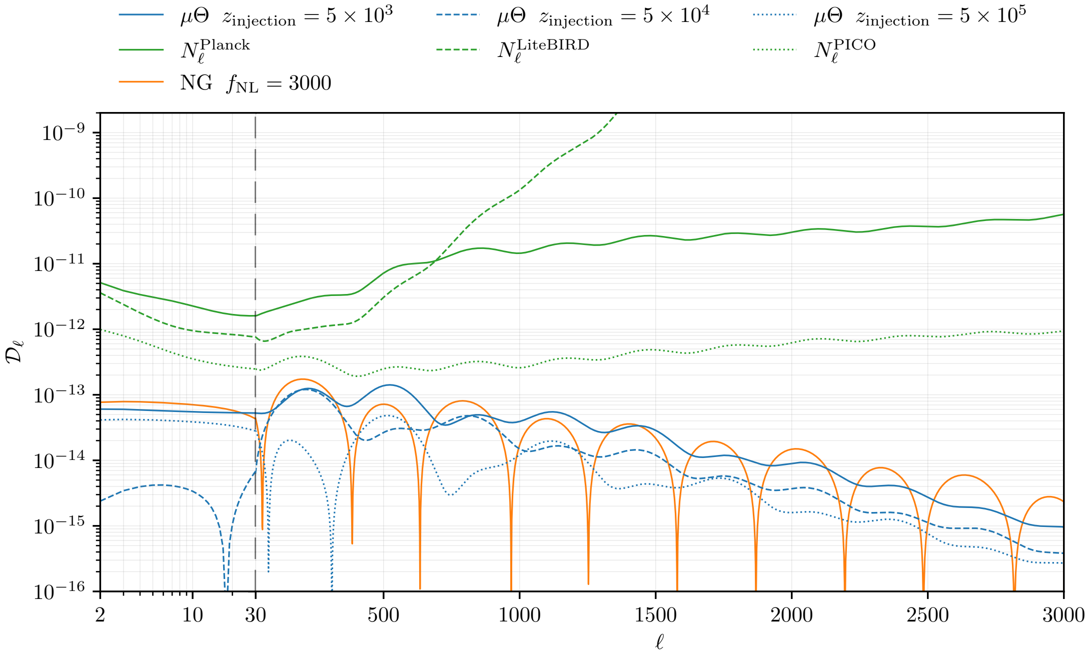


Figure 26: A figure showing the expected $\Theta \times \mu$ power spectra (blue) for various injection times with energy release $\Delta\rho/\rho = 3 \times 10^{-5}$. Also shown for comparison are the noise curves (green) for various instruments and a predicted signal from primordial non-gaussianity (orange).

The same exercise is repeated in Fig. 27 for the $\Theta{\times}y$ cross correlation. The only difference is that here we show for reference (orange) the Sunyaev-Zeldovich (SZ) cross correlation with ISW, tSZ×Θ. This signal would in principle constitute a bias to the $\Theta \times y$ cross correlation from energy injection. Here we disregard this problem; however, we point out that the vastly different ℓ dependence would possibly allow for a successful signal disentanglement. Conversely, existing primordial distortion anisotropies would provide a noise contribution to SZ searches for the ISW effect [67, 68]. We also specify that this contribution is included in the covariance calculation, but as one can expect, it has a negligible effect on the results.

In Fig. 28 we show the constraints on single injection scenario as a function of the injection redshift. In all the panels we can appreciate a subtle distinctions in three regimes which coincide with the standard y, residual and μ-eras. Generally speaking μ-era injection is less constrainable than the residual and y-era injections. In Fig. 6 we see that the perturbed thermalisation and anisotropic heating sources actually oppose the boosting source for early injection times. The boosting source however flips the sign of its $\mu^{(1)}$ source as the background spectrum contains more contributions of $y^{(0)}$. In Fig. 9 this leads to an additive effect of boosting for late times. This likely explains both the lack of constraining power at early times as well as the small step around $z \approx 7 \times 10^4$ within each panel of Fig. 28. The other small step occurs around $z \approx 2 \times 10^5$, hinting towards the thermalisation terms becoming inefficient. Similarly a small decrease of constraining power is seen at $z \approx 10^6$ since part of the distortion thermalised to a simple temperature shift.

Combining constraints from μ and y distortions would allow us to set tight limits on the energy injection throughout the whole post-T-era universe history. In particular next generation and futuristic satellites, thanks to their ability to remove foregrounds due to ample frequency coverage, will set constraints exceeding *COBE/FIRAS*'. Further to μ and y we could feasibly use the residual r distor-

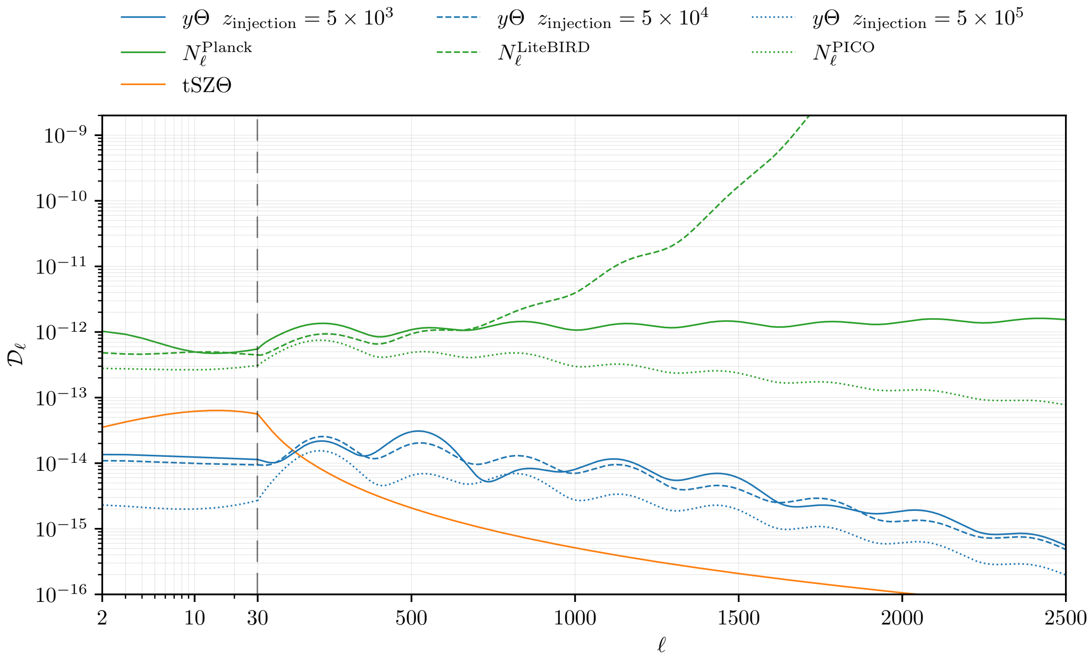


Figure 27: A figure showing the expected $\Theta \times y$ power spectra (blue) for various injection times with energy release $\Delta\rho/\rho = 3 \times 10^{-5}$. Also shown for comparison are the noise curves (green) for various instruments and a predicted signal from the thermal Sunyaev-Zeldovich effect (orange).

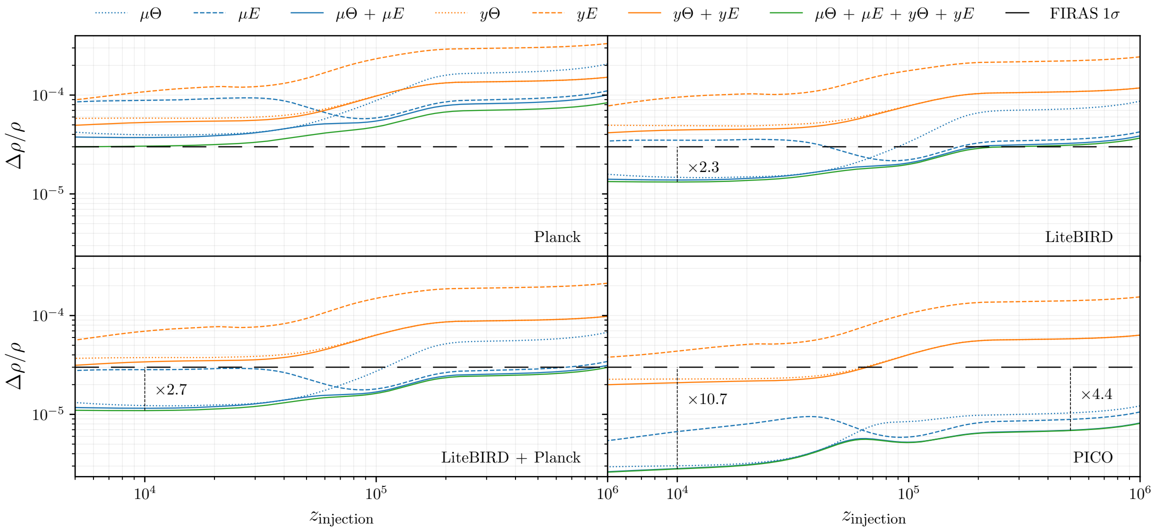


Figure 28: Four figures showing the forecasts for time dependent constraints on single energy injection using different cross correlations. Also shown is the *COBE/FIRAS* limit on energy release at a time-independent $\Delta\rho/\rho = 3 \times 10^{-5}$. How much a given instrument could improve on the *COBE/FIRAS* measurement is shown in multiplicative factors in the plot.

tions to improve the results further. These however are at least an order of magnitude smaller as seen in Sect. 4, but could carry details of time dependence. We will carry on a more detailed discussion in the next section.

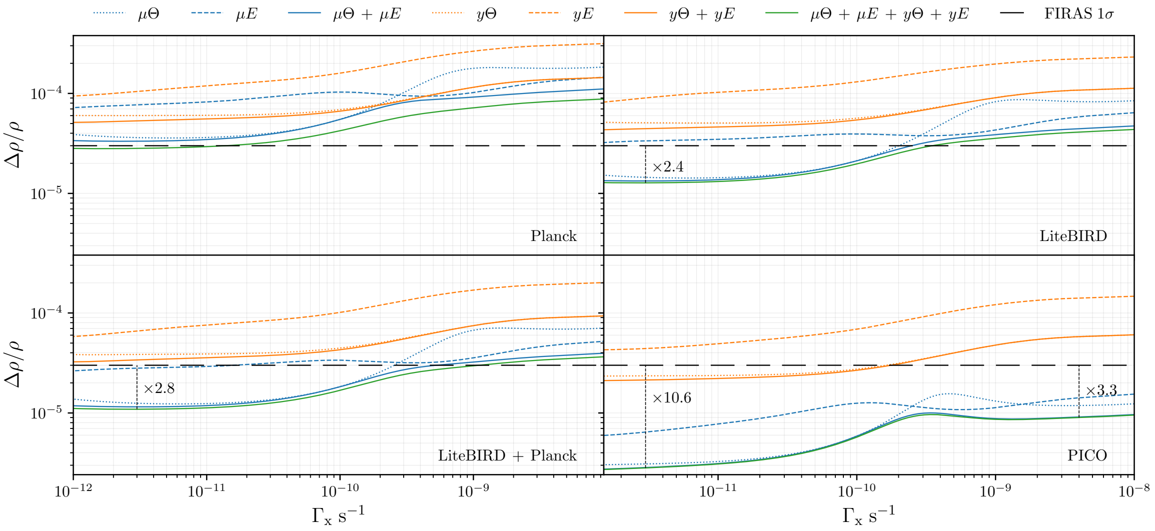


Figure 29: Four figures showing the forecasts for decaying particle constraints using different cross correlations. Crucially these constraints ignore the effects of perturbed decay. Also shown is the FIRAS limit on energy release at a time-independent $\Delta\rho/\rho = 3 \times 10^{-5}$. How much a given instrument could improve on the *COBE/FIRAS* measurement is shown in multiplicative factors in the plot.

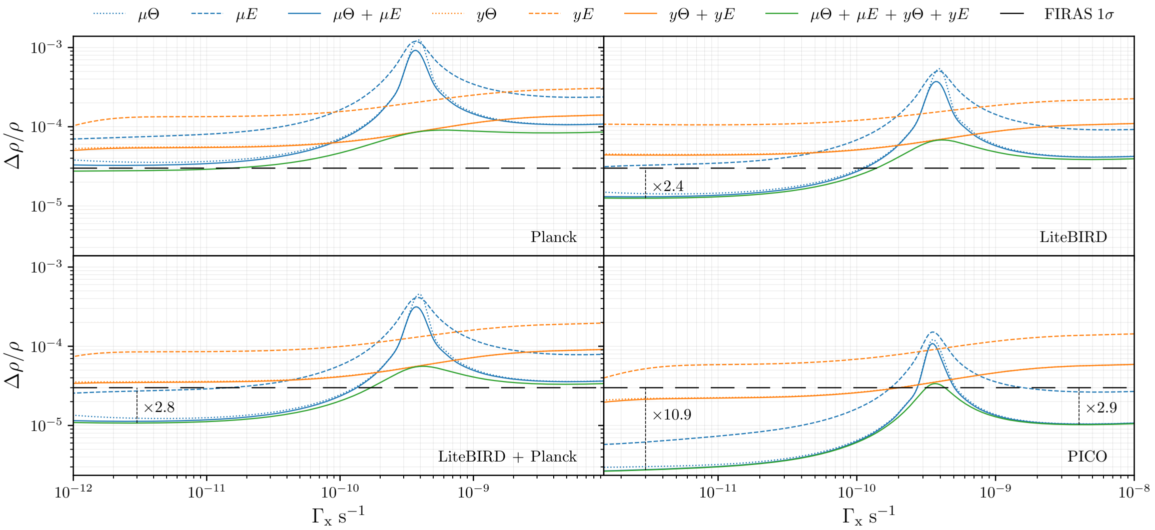


Figure 30: As for Fig. 29, but now including the effects of perturbed decay. The constraints are reduced around the residual distortion era.

In Fig. 29 and Fig. 30 we show the equivalent constraints for decaying particle scenarios of different lifetimes with and without perturbed decay. As seen above from directly inspecting the power spectra, these models hold many similarities with the single-injection scenarios. The big differences emerge when including the effects of perturbed decay. Interestingly these serve to decrease constraining power, which may initially be counter intuitive. One problem for constraining this model is that the enhancements from dark matter modulations typically occur for $\ell > 1000$ (see Fig. 25) while the maximum constraining power usually comes from $100 \leq \ell \leq 1000$. Furthermore we previously commented that the combination of $\Psi^{(1)} + \delta^{(1)}_{\rm cdm}$ in adiabatic initial conditions imply some mutual cancellation, thus reducing the overall effect of anisotropic heating. With anisotropic heating effectively halved (and flipped sign) the early time constraints decrease, and a noticeable peak emerges. This peak is due to the dependence on boosting sources, which between early and late times cross a specific mix of $y^{(0)}$ and $\mu^{(0)}$ which boosts to give no $\mu^{(1)}$ contribution.

This example is illustrative of the high degree of model dependence in constraints when it comes to energy injection modulated directed by local perturbed quantities. On the other hand, the case without perturbed decay illustrates that single injection models do a good job of representing continuous injection mechanisms assuming they have narrow windows. Overall, our forecasts demonstrate the immense potential for SD anisotropy studies with CMB imagers.

5.1 Accessing information from the residual distortions

As we have seen in Sect. 4, additional information could be gleaned from the residual distortion signals. In the estimates given above, we neglected this component for several reasons. Firstly, one does not expect this to improve the detection limits by more than $\simeq 10-20\%$, given that the overall energy contained in the residual spectra is at that level relative to the y-distortion. However, in the residual distortion era, the effect could be slightly more noticeable. Secondly, the residual distortion modes used in our presentation where constructed for a *PIXIE*-like configuration [69, 70]. As such, one cannot directly translate these to the experimental configurations assumed in the forecasts, and an experiment-specific analysis would be required. Thirdly, for μ and y-type distortion signals, we already have some level of understanding about how foregrounds might affect the constraints [56, 58, 71], but for the residual distortion modes, a more comprehensive analysis is outstanding. Solving these issues is beyond the scope of this paper, such that we do not attempt further improvements of our forecasts. We note, however, that in particular in the residual distortion era, one expects additional gains. In addition, should one detect a SD signal with future CMB imagers, we expect the residual distortion information to provide more leverage towards distinguishing various distortion signals, as also understood from the average distortion science [32, 40].

5.2 Further optimizing and reducing the observation basis

We close by remarking on further optimizing the observation basis. The construction so far was motivated by conserving the meaning of the standard CMB signals, G, Y and M. However, we now understand that for the SD anisotropies the $\hat{O}_x Y$, $\hat{O}_x M$, $\hat{\mathcal{D}}_x M$ and $Y - Y_1$ SEDs also play central roles.

A pure y-anisotropy can be created by late perturbed energy release ($z \lesssim 10^4$), as in the example of decaying particles. Another example would be perturbed dark matter annihilation, SZ clusters or the dissipation of acoustic modes from primordial perturbations, which all intrinsically source Y. Pure μ-distortion anisotropies can be sourced by perturbed energy release at $z \gtrsim 10^5$ from annihilation/decay or acoustic mode dissipation.

Without imposing a theory prior, one should therefore consider adding $\hat{O}_x Y = 4Y_1$, $\hat{O}_x M$ and $\hat{\mathcal{D}}_x M$ to the standard distortion shapes relevant to CMB anisotropy analysis. This means, in total six CMB SED amplitudes would have to be determined. Given that $\hat{O}_x M$ is extremely well represented by a simple superposition of Y and M, large degeneracies would be found. Similarly, $\hat{O}_x Y$ does have significant projections onto G, Y and M, such that without further study it is unclear how much this extended standard SED basis would help.

However, if one can with certainty assume that the energy release occurred at $z \lesssim 10^4$, our computations have demonstrated (see Fig. 3) that one could indeed simply use G, Y and $\hat{O}_x Y = 4Y_1$ for the full analysis. This would ensure that no information is lost, even if the SEDs are not fully independent. In reality, in this regime, one could even get away with simply using a Y_1-distortion

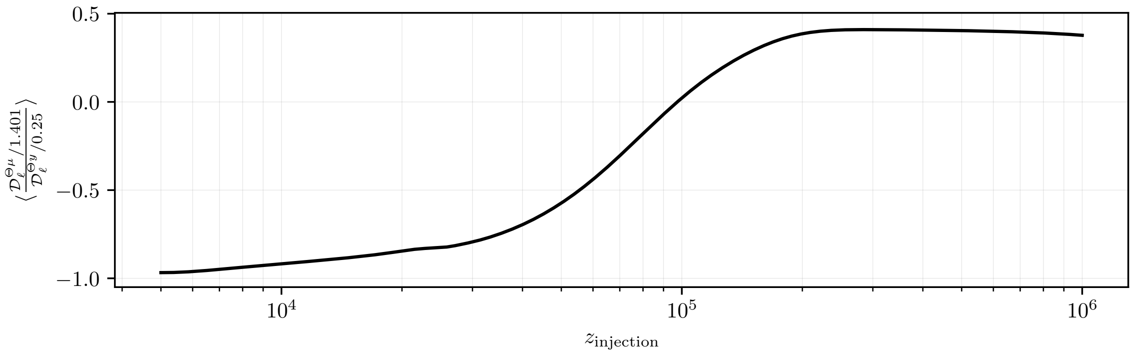


Figure 31: A figure showing the average ratio between $\mathcal{D}_\ell^{\Theta\mu}$ and $\mathcal{D}_\ell^{\Theta y}$ for different single injection redshifts with appropriately normalised energy units. The average is taken in the range $100 \le \ell \le 1000$ where most of the constraining power is.

hierarchy without the need of further SED rotation. We will investigate the utility of these alternative analysis methods in the future, highlighting that the current observation basis provides a more agnostic approach to the challenge of extracting all the information from the CMB sky.

5.3 Extracting the time of injection

In much of Sect. 4 we alluded to strong time dependence encoded in the power spectrum. Consider for example Fig. 19 where by taking ratios of the measured CMB power spectra there is a relatively smooth gradient between early and late injection times. To understand what time sensitivity future imagers might have on energy injection times we take an average of this ratio in the range $100 \le \ell \le 1000$, where we have seen most of the constraining power to reside. This process is shown in Fig. 31, where we can see the previously discussed plateau for early and late injection time, as well as a strong gradient for residual-era injection. Changing the energy release model subtly modifies the shape of the curve, but the general trends within each era hold regardless of exact energy injection mechanism.

The figure simply shows that constrainable metrics exist for time dependence, however, a much more robust approach would be to perform full MCMC searches for a given models parameters. This is motivated by the fact that individual peaks of the spectra could break the plateaus deep into a given era (e.g. the first and second peak in Fig. 18 would reveal how late into the y-era energy injection occurred). To perform these kind of searches it would be necessary to optimise the numerical treatment. With future work this should be possible, considering that the largest burden in the current code is pushing to $N_{\rm max} = 15$, which as seen in Fig. 32 and Fig. 33 yields small percent changes which could be neglected for broad parameter searches.

6 Discussion and Conclusions

This paper has taken a step-by-step approach to presenting the spectro-spatial evolution of the photon spectrum. Starting with the photon spectrum, we discussed dominant contributions to the anisotropic spectrum and the different limits they see their greatest importance. We followed with a careful treatment of basis choices for representing these spectra as transfer functions of SED amplitudes. This opened a long discussion on the relevant physics for different injection eras, which culminated in the presentation of distortion power spectra, which show a complex superposition of many physical effects previously analysed. Finally, we used the predicted power spectra to forecast constraints on primordial energy injection which are comparable with or exceed modern limits from *COBE/FIRAS*, thus constituting independent tests of primordial physics.

This detailed step-by-step approach should not distract from the novelty of this formalism and the wealth of opportunities in it: it is now possible to infer valuable information about the background photon spectrum in a potentially more time sensitive way by viewing the spectrum in anisotropic patches of the CMB sky. This allows us to push our understanding of primordial physics beyond the opaque *CMB curtain* and gain insight to $z \simeq 2 \times 10^6$ using well-known CMB imaging techniques.

In essence the time sensitivity arises from the fact that we can distinguish anisotropic heating [creating $M(x)$, $Y(x)$ or $Y_1 + Y(x)$], perturbed thermalisation [creating $\hat{\mathcal{D}}_x M(x)$ and $Y_1(x) - Y(x)$] and boosting [producing $\hat{O}_x M(x)$ or $\hat{O}_x Y(x)$ depending on the background spectrum]. Fully exploiting this time dependence will require more sophisticated analysis tools and a further optimised numerics (see Sect. 5.3). The variety of spectral modes has strong potential for discriminating and isolating the physical origin of any energy injection in the primordial plasma, even if it is clear that additional optimisation of the analysis may be needed (see Sect. 5.2).

In our analysis we clearly isolated the main effects: Early energy injection see source terms arise from Compton scattering, perturbed scattering, perturbed emission and direct anisotropic heating. At late times the main sources are Doppler and potential driving, which are well-known in connection with the acoustic peaks seen in the regular temperature power spectra. Each of these were individually illustrated in detail, showing that new insight can be gained deep into the pre-recombination era (e.g., see Figs. 9, 15 and 20) by accessing the full spectro-spatial information.

The probably most significant conclusion from this work is that through measurements of SD anisotropies CMB imagers like *Planck* [59], *LiteBIRD* [60] and *PICO* [61] can provide limits on the *average energy release* at various epochs. These limits are expected to be comparable with those from *COBE/FIRAS* in the case of *Planck*, while *LiteBIRD* and *PICO* could supersede *COBE/FIRAS* several times (see Sect. 5). For *PICO*, we see possible gains by more than one order of magnitude for energy injection at $z \lesssim 5 \times 10^4$ (see Fig. 28). Evidently, a *PIXIE*-like CMB spectrometer, directly targeting the average distortion, could improve these limits many-fold [e.g., 72, 73]; however, the experimental methods and analysis techniques are quite different, and the complementarity of the constraints highlights the unique synergy between the approaches.

Although our estimates are based on simple Fisher forecasts, similar methods have already been shown to be reliable [58], building confidence in our results. Nevertheless, we plan to carry out a more rigorous analysis using detailed foreground simulations and including other experimental effects to further refine these results in future work. We do not expect the main conclusion to change: CMB imaging and distortion science have now been united to allow full spectro-spatial considerations of CMB physics. Extended forecasts that consider the benefits of ground-based observations with The Simons Observatory [74], CMB-S4 [75] and other space-mission concepts [76–79] are also planned.

We close with a few words about the path forward. First and foremost, it would be important to improve the analytic understanding of the solutions. In addition, our solutions clearly show that in the tight-coupling regime the distortion dipole transfer functions all follow $y_{n,1} \propto \mu_1 \propto \Theta_1$ once the average distortion has been created (see Sect. 3). With this tracking solution and approximation for the potential Φ, the distortion monopoles can be modelled using WKB approximations. Some complications arise from the SD evolution by Compton scattering terms and the transformation to the optimal observation basis. However, we leave a more detailed study to future work.

A simple and immediate extension of this work is to investigate the effects of different initial conditions beyond adiabatic modes as well as the effects of different heating mechanisms. In this paper the case of perturbed decay as a trace of dark matter illustrated the importance of both these avenues of exploration. Firstly the adiabaticity of the initial perturbations are capable of cancelling or enhancing important thermalisation terms, providing an additional test to the usual CMB temperature power spectra. A comprehensive study in this direction would complement the work on initial conditions and SD physics for the average photon spectrum [80]. Secondly, modulating local heating by perturbed quantities with known effects on the temperature power spectra yield powerful and predictable correlations. If observable, these concrete enhancements to the CMB power spectra could provide *smoking guns* of concrete heating mechanisms and therefore new physics.

Furthermore, here we only illustrated the SD anisotropy physics for single energy injection and decaying particle scenarios, focusing on the distortion signals. However, other mechanisms can be considered. For instance, dark matter annihilation should similarly lead to anisotropic distortions. For s-wave (i.e., temperature-independent) annihilation, these signals ought to be small from the pre-recombination era, given existing constraints from *Planck* [e.g., 2]; however, due to perturbed decay, $\propto \delta_{\rm cdm}$, late time y-type anisotropies could be sourced by the non-linear growth of structure, enhancing the expected signal. Sommerfeld boosts of the annihilation rate [81] or a varying temperature dependence of the annihilation cross section [e.g., p-wave annihilation 40] could further modify the signals. In addition, cosmic bubble collisions [82–84], primordial black holes [85–88], primordial magnetic fields [89–92] or dark matter scattering effects [93–95] could lead to anisotropic energy release, which can now be modelled more accurately using our novel approach. For this, the possible changes to the ionisation history [44, 45, 96, 97] should be taken into account using state of the art recombination codes like `CosmoRec`, an extension that we plan for the near future.

The computations of the SD anisotropies from primordial non-Gaussianity [65, 98] could also be refined, accurately treating all the transfer effects [see 99, 100, for some previous analytic attempts], which will be crucial for distinguishing these signals from contaminations and foregrounds. Specifically, this latter problem deserves additional attention, as the SD anisotropies we considered here could act as a new foreground to extracting information about primordial non-Gaussianity as long as we have no significantly improved upper limit on average energy release from absolute spectrometers such as *PIXIE*. This can be appreciated from Fig. 26, which demonstrates that at the level of $f_{\rm nl} = 3000$, the possible anisotropic distortion signals due to average energy release are comparable. Without theoretical prior, it will be challenging to eliminate this uncertain contribution using CMB imaging alone. Even if the $\mu \times T$ signal from primordial non-Gaussianity differs from the signals discussed here, it will be hard to reach a cosmic-variance limited measurement suggested by theory [98, 101] unless we could limit the average energy release to $\Delta\rho/\rho \lesssim 10^{-11}$. Similar comments could impede polarisation-distortion correlation studies, which have the potential to shed new light on inflation physics [102]. Attempts to extract information about the ISW effect from SZ cluster-induced $y \times T$ correlations [67, 68] would also be affected. We plan to investigate these aspect more carefully in future work, hoping that the new perspectives given here provide further motivation to think about an extended synergistic approach in the future of CMB exploration.

Acknowledgments

We thank Eiichiro Komatsu, Aditya Rotti and Rashid Sunyaev for stimulating discussion. We furthermore thank Colin Hill, Rishi Khatri, Anthony Lewis, Atsuhisa Ota, Enrico Pajer and Nils Schöneberg for comments on the manuscript. This work was supported by the ERC Consolidator Grant *CMBSPEC* (No. 725456). TK was also supported by STFC grant ST/T506291/1. JC was furthermore supported by the Royal Society as a Royal Society University Research Fellow at the University of Manchester, UK (No. URF/R/191023). AR acknowledges support by the project "Combining Cosmic Microwave Background and Large Scale Structure data: an Integrated Approach for Addressing Fundamental Questions in Cosmology", funded by the MIUR Progetti di Ricerca di Rilevante Interesse Nazionale (PRIN) Bando 2017 - grant 2017YJYZAH.

References

[1] C. L. Bennett, M. Halpern, G. Hinshaw, et al., *First-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Preliminary Maps and Basic Results*, *The Astrophysical Journal Supplement* **148** (Sept., 2003) 1–27.

[2] Planck Collaboration, P. A. R. Ade, N. Aghanim, et al., *Planck 2015 results. XIII. Cosmological parameters*, *Astronomy & Astrophysics* **594** (Sept., 2016) A13, [`arXiv:1502.01589`].

[3] R. A. Sunyaev and Y. B. Zeldovich, *Small-Scale Fluctuations of Relic Radiation*, *Astrophysics and Space Science* **7** (1970) 3–+.

[4] P. J. E. Peebles and J. T. Yu, *Primeval Adiabatic Perturbation in an Expanding Universe*, *The Astrophysical Journal* **162** (Dec., 1970) 815–+.

[5] J. Silk, *Cosmic Black-Body Radiation and Galaxy Formation*, *The Astrophysical Journal* **151** (Feb., 1968) 459–+.

[6] W. Hu and N. Sugiyama, *Anisotropies in the cosmic microwave background: an analytic approach*, *The Astrophysical Journal* **444** (May, 1995) 489–506, [`astro-ph/9407093`].

[7] C.-P. Ma and E. Bertschinger, *Cosmological Perturbation Theory in the Synchronous and Conformal Newtonian Gauges*, *The Astrophysical Journal* **455** (Dec., 1995) 7–+, [`astro-ph/9506072`].

[8] U. Seljak and M. Zaldarriaga, *Signature of Gravity Waves in the Polarization of the Microwave Background*, *Physical Review Letters* **78** (Mar., 1997) 2054–2057, [`astro-ph/9609169`].

[9] W. Hu and M. White, *CMB anisotropies: Total angular momentum method*, *Physical Review D* **56** (July, 1997) 596–615, [`astro-ph/9702170`].

[10] Y. B. Zeldovich and R. A. Sunyaev, *The Interaction of Matter and Radiation in a Hot-Model Universe*, *Astrophysics and Space Science* **4** (July, 1969) 301–316.

[11] R. A. Sunyaev and Y. B. Zeldovich, *The interaction of matter and radiation in the hot model of the Universe, II*, *Astrophysics and Space Science* **7** (Apr., 1970) 20–30.

[12] A. F. Illarionov and R. A. Sunyaev, *Comptonization, characteristic radiation spectra, and thermal balance of low-density plasma*, *Soviet Astronomy* **18** (Feb., 1975) 413–419.

[13] A. F. Illarionov and R. A. Sunyaev, *Comptonization, the background-radiation spectrum, and the thermal history of the universe*, *Soviet Astronomy* **18** (June, 1975) 691–699.

[14] L. Danese and G. de Zotti, *Double Compton process and the spectrum of the microwave background*, *Astronomy & Astrophysics* **107** (Mar., 1982) 39–42.

[15] C. Burigana, *Distortions of the CMB Spectrum by Continuous Heating*, in *Observational Cosmology* (G. L. Chincarini, A. Iovino, T. Maccacaro, & D. Maccagni, ed.), vol. 51 of *Astronomical Society of the Pacific Conference Series*, pp. 554–+, Jan., 1993.

[16] W. Hu and J. Silk, *Thermalization and spectral distortions of the cosmic background radiation*, *Physical Review D* **48** (July, 1993) 485–502.

[17] J. Chluba, T. Kite, and A. Ravenni, *Spectro-spatial evolution of the CMB I: discretisation of the thermalisation Green's function*, *arXiv e-prints* (Oct., 2022) arXiv:2210.09327, [`arXiv:2210.09327`].

[18] J. Chluba, A. Ravenni, and T. Kite, *Spectro-spatial evolution of the CMB II: generalised Boltzmann hierarchy*, *arXiv e-prints* (Oct., 2022) arXiv:2210.15308, [`arXiv:2210.15308`].

[19] J. Chluba, *Green's function of the cosmological thermalization problem*, *Monthly Notices of the Royal Astronomical Society* **434** (Sept., 2013) 352–357, [`arXiv:1304.6120`].

[20] M. Lucca, N. Schöneberg, D. C. Hooper, J. Lesgourgues, and J. Chluba, *The synergy between CMB spectral distortions and anisotropies*, *Journal of Cosmology and Astroparticle Physics* **2020** (Feb., 2020) 026, [`arXiv:1910.04619`].

[21] J. Chluba, R. Khatri, and R. A. Sunyaev, *CMB at 2 x 2 order: the dissipation of primordial acoustic waves and the observable part of the associated energy release*, *Monthly Notices of the Royal Astronomical Society* **425** (Sept., 2012) 1129–1169, [`arXiv:1202.0057`].

[22] L. Danese and G. de Zotti, *The relic radiation spectrum and the thermal history of the Universe*, *Nuovo Cimento Rivista Serie* **7** (Sept., 1977) 277–362.

[23] S. A. Balashev, E. E. Kholupenko, J. Chluba, A. V. Ivanchik, and D. A. Varshalovich, *Spectral Distortions of the CMB Dipole*, *The Astrophysical Journal* **810** (Sept., 2015) 131, [`arXiv:1505.06028`].

[24] G. De Zotti, M. Negrello, G. Castex, A. Lapi, and M. Bonato, *Another look at distortions of the Cosmic Microwave Background spectrum*, *Journal of Cosmology and Astroparticle Physics* **3** (Mar., 2016) 047, [`arXiv:1512.04816`].

[25] J. Chluba and R. A. Sunyaev, *The evolution of CMB spectral distortions in the early Universe*, *Monthly Notices of the Royal Astronomical Society* **419** (Jan., 2012) 1294–1314, [`arXiv:1109.6552`].

[26] S. K. Acharya and J. Chluba, *CMB spectral distortions from continuous large energy release*, *arXiv e-prints* (Dec., 2021) arXiv:2112.06699, [`arXiv:2112.06699`].

[27] A. Lewis, A. Challinor, and A. Lasenby, *Efficient Computation of Cosmic Microwave Background Anisotropies in Closed Friedmann-Robertson-Walker Models*, *The Astrophysical Journal* **538** (Aug., 2000) 473–476, [`astro-ph/9911177`].

[28] J. Lesgourgues, *The Cosmic Linear Anisotropy Solving System (CLASS) I: Overview*, *ArXiv:1104.2932* (Apr., 2011) [`arXiv:1104.2932`].

[29] J. Chluba, G. M. Vasil, and L. J. Dursi, *Recombinations to the Rydberg states of hydrogen and their effect during the cosmological recombination epoch*, *Monthly Notices of the Royal Astronomical Society* **407** (Sept., 2010) 599–612, [`arXiv:1003.4928`].

[30] J. Chluba, *Refined approximations for the distortion visibility function and μ-type spectral distortions*, *Monthly Notices of the Royal Astronomical Society* **440** (Apr., 2014) 2544–2563, [`arXiv:1312.6030`].

[31] J. Chluba, *Green's function of the cosmological thermalization problem - II. Effect of photon injection and constraints*, *Monthly Notices of the Royal Astronomical Society* **454** (Dec., 2015) 4182–4196, [`arXiv:1506.06582`].

[32] J. Chluba and D. Jeong, *Teasing bits of information out of the CMB energy spectrum*, *Monthly Notices of the Royal Astronomical Society* **438** (Mar., 2014) 2065–2082, [`arXiv:1306.5751`].

[33] J. Chluba and R. M. Thomas, *Towards a complete treatment of the cosmological recombination problem*, *Monthly Notices of the Royal Astronomical Society* **412** (Apr., 2011) 748–764, [`arXiv:1010.3631`].

[34] C. Burigana, L. Danese, and G. de Zotti, *Formation and evolution of early distortions of the microwave background spectrum - A numerical study*, *Astronomy & Astrophysics* **246** (June, 1991) 49–58.

[35] L. Senatore, S. Tassev, and M. Zaldarriaga, *Non-gaussianities from perturbing recombination*, *Journal of Cosmology and Astroparticle Physics* **9** (Sept., 2009) 38, [`arXiv:0812.3658`].

[36] D. Zegeye, K. Inomata, and W. Hu, *Spectral distortion anisotropy from inflation for primordial black holes*, *Physical Review D* **105** (May, 2022) 103535, [`arXiv:2112.05190`].

[37] J. Chluba, *Tests of the CMB temperature-redshift relation, CMB spectral distortions and why adiabatic photon production is hard*, *Monthly Notices of the Royal Astronomical Society* **443** (Sept., 2014) 1881–1888, [`arXiv:1405.1277`].

[38] G. Steigman, *Primordial Nucleosynthesis in the Precision Cosmology Era*, *Annual Review of Nuclear and Particle Science* **57** (Nov., 2007) 463–491, [arXiv:0712.1100].

[39] G. Steigman, *Tracking the post-bbn evolution of deuterium*, *arXiv* **astro-ph.GA** (Jan, 2009) [arXiv:0901.4333v].

[40] J. Chluba, *Distinguishing different scenarios of early energy release with spectral distortions of the cosmic microwave background*, *Monthly Notices of the Royal Astronomical Society* **436** (Dec., 2013) 2232–2243, [arXiv:1304.6121].

[41] S. Sarkar and A. M. Cooper, *Cosmological and experimental constraints on the tau neutrino*, *Physics Letters B* **148** (Nov., 1984) 347–354.

[42] W. Hu and J. Silk, *Thermalization constraints and spectral distortions for massive unstable relic particles*, *Physical Review Letters* **70** (May, 1993) 2661–2664.

[43] B. Bolliet, J. Chluba, and R. Battye, *Spectral distortion constraints on photon injection from low-mass decaying particles*, *arXiv e-prints* (Dec., 2020) arXiv:2012.07292, [arXiv:2012.07292].

[44] X. Chen and M. Kamionkowski, *Particle decays during the cosmic dark ages*, *Physical Review D* **70** (Aug., 2004) 043502–+, [astro-ph/0310473].

[45] N. Padmanabhan and D. P. Finkbeiner, *Detecting dark matter annihilation with CMB polarization: Signatures and experimental prospects*, *Physical Review D* **72** (July, 2005) 023508–+, [astro-ph/0503486].

[46] P. Callin, *How to calculate the CMB spectrum*, *arXiv e-prints* (June, 2006) astro–ph/0606683, [astro-ph/0606683].

[47] D. J. Fixsen, E. S. Cheng, J. M. Gales, et al., *The Cosmic Microwave Background Spectrum from the Full COBE FIRAS Data Set*, *The Astrophysical Journal* **473** (Dec., 1996) 576–+, [astro-ph/9605054].

[48] D. J. Fixsen, *The Temperature of the Cosmic Microwave Background*, *The Astrophysical Journal* **707** (Dec., 2009) 916–920, [arXiv:0911.1955].

[49] M. Remazeilles, J. Delabrouille, and J.-F. Cardoso, *CMB and SZ effect separation with Constrained Internal Linear Combinations*, *Mon. Not. Roy. Astron. Soc.* **410** (2011) 2481, [arXiv:1006.5599].

[50] A. Cooray, W. Hu, and M. Tegmark, *Large scale Sunyaev-Zel'dovich effect: Measuring statistical properties with multifrequency maps*, *Astrophys. J.* **540** (2000) 1–13, [astro-ph/0002238].

[51] J. C. Hill and E. Pajer, *Cosmology from the thermal Sunyaev-Zel'dovich power spectrum: Primordial non-Gaussianity and massive neutrinos*, *Phys. Rev.* **D88** (2013), no. 6 063526, [arXiv:1303.4726].

[52] A. Ravenni, M. Rizzato, S. Radinović, et al., *Breaking degeneracies with the Sunyaev-Zeldovich full bispectrum*, *JCAP* **06** (2021) 026, [arXiv:2008.12947].

[53] M. H. Abitbol, J. Chluba, J. C. Hill, and B. R. Johnson, *Prospects for Measuring Cosmic Microwave Background Spectral Distortions in the Presence of Foregrounds*, *Mon. Not. Roy. Astron. Soc.* **471** (2017), no. 1 1126–1140, [arXiv:1705.01534].

[54] M. Tegmark, D. J. Eisenstein, W. Hu, and A. de Oliveira-Costa, *Foregrounds and forecasts for the cosmic microwave background*, *Astrophys. J.* **530** (2000) 133–165, [astro-ph/9905257].

[55] J. Dunkley et al., *The Atacama Cosmology Telescope: likelihood for small-scale CMB data*, *JCAP* **07** (2013) 025, [arXiv:1301.0776].

[56] M. Remazeilles, A. Ravenni, and J. Chluba, *Leverage on small-scale primordial non-Gaussianity through cross-correlations between CMB E-mode and μ-distortion anisotropies*, *Monthly Notices of the Royal Astronomical Society* **512** (May, 2022) 455–470, [arXiv:2110.14664].

[57] M. Remazeilles and J. Chluba, *Extracting foreground-obscured μ-distortion anisotropies to constrain primordial non-Gaussianity*, *Mon. Not. Roy. Astron. Soc.* **478** (2018), no. 1 807–824, [arXiv:1802.10101].

[58] A. Rotti, A. Ravenni, and J. Chluba, *Non-Gaussianity constraints from Planck spectral distortion cross-correlations*, *arXiv e-prints* (May, 2022) arXiv:2205.15971, [`arXiv:2205.15971`].

[59] **Planck** Collaboration, N. Aghanim et al., *Planck 2018 results. I. Overview and the cosmological legacy of Planck*, *Astron. Astrophys.* **641** (2020) A1, [`arXiv:1807.06205`].

[60] **LiteBIRD** Collaboration, M. Hazumi et al., *LiteBIRD: JAXA's new strategic L-class mission for all-sky surveys of cosmic microwave background polarization*, *Proc. SPIE Int. Soc. Opt. Eng.* **11443** (2020) 114432F, [`arXiv:2101.12449`].

[61] **NASA PICO** Collaboration, S. Hanany et al., *PICO: Probe of Inflation and Cosmic Origins*, `arXiv:1902.10541`.

[62] J. Dick, M. Remazeilles, and J. Delabrouille, *Impact of calibration errors on CMB component separation using FastICA and ILC*, *Monthly Notices of the Royal Astronomical Society* **401** (Jan., 2010) 1602–1612, [`arXiv:0907.3105`].

[63] J. Chluba, *Which spectral distortions does ΛCDM actually predict?*, *Monthly Notices of the Royal Astronomical Society* **460** (July, 2016) 227–239, [`arXiv:1603.02496`].

[64] A. Lewis, *Rayleigh scattering: blue sky thinking for future CMB observations*, *Journal of Cosmology and Astroparticle Physics* **2013** (Aug., 2013) 053, [`arXiv:1307.8148`].

[65] J. Ganc and E. Komatsu, *Scale-dependent bias of galaxies and μ-type distortion of the cosmic microwave background spectrum from single-field inflation with a modified initial state*, *Physical Review D* **86** (July, 2012) 023518, [`arXiv:1204.4241`].

[66] A. Ravenni, M. Liguori, N. Bartolo, and M. Shiraishi, *Primordial non-Gaussianity with μ-type and y-type spectral distortions: exploiting Cosmic Microwave Background polarization and dealing with secondary sources*, *Journal of Cosmology and Astroparticle Physics* **2017** (Sept., 2017) 042, [`arXiv:1707.04759`].

[67] N. Taburet, C. Hernández-Monteagudo, N. Aghanim, M. Douspis, and R. A. Sunyaev, *The ISW-tSZ cross-correlation: integrated Sachs-Wolfe extraction out of pure cosmic microwave background data*, *Monthly Notices of the Royal Astronomical Society* **418** (Dec., 2011) 2207–2218, [`arXiv:1012.5036`].

[68] C. Creque-Sarbinowski, S. Bird, and M. Kamionkowski, *Cross-correlation between thermal Sunyaev-Zeldovich effect and the integrated Sachs-Wolfe effect*, *Physical Review D* **94** (Sept., 2016) 063519, [`arXiv:1606.00839`].

[69] A. Kogut, J. Chluba, D. J. Fixsen, S. Meyer, and D. Spergel, *The Primordial Inflation Explorer (PIXIE)*, in *SPIE Conference Series*, vol. 9904 of *Proc.SPIE*, p. 99040W, July, 2016.

[70] A. Kogut, M. H. Abitbol, J. Chluba, et al., *CMB Spectral Distortions: Status and Prospects*, in *Bulletin of the American Astronomical Society*, vol. 51, p. 113, Sept., 2019. `arXiv:1907.13195`.

[71] M. H. Abitbol, J. Chluba, J. C. Hill, and B. R. Johnson, *Prospects for Measuring Cosmic Microwave Background Spectral Distortions in the Presence of Foregrounds*, *Monthly Notices of the Royal Astronomical Society* (May, 2017) [`arXiv:1705.01534`].

[72] J. Chluba, A. Kogut, S. P. Patil, et al., *Spectral Distortions of the CMB as a Probe of Inflation, Recombination, Structure Formation and Particle Physics*, *Bulletin of the American Astronomical Society* **51** (May, 2019) 184, [`arXiv:1903.04218`].

[73] J. Chluba, M. H. Abitbol, N. Aghanim, et al., *New horizons in cosmology with spectral distortions of the cosmic microwave background*, *Experimental Astronomy* **51** (June, 2021) 1515–1554, [`arXiv:1909.01593`].

[74] The Simons Observatory Collaboration, P. Ade, J. Aguirre, et al., *The Simons Observatory: Science goals and forecasts*, *ArXiv:1808.07445* (Aug., 2018) [`arXiv:1808.07445`].

[75] K. N. Abazajian, P. Adshead, Z. Ahmed, et al., *CMB-S4 Science Book, First Edition*, *ArXiv:1610.02743* (Oct., 2016) [`arXiv:1610.02743`].

[76] PRISM Collaboration, P. Andre, C. Baccigalupi, et al., *PRISM (Polarized Radiation Imaging and Spectroscopy Mission): A White Paper on the Ultimate Polarimetric Spectro-Imaging of the Microwave and Far-Infrared Sky*, *ArXiv:1306.2259* (June, 2013) [`arXiv:1306.2259`].

[77] P. André, C. Baccigalupi, A. Banday, et al., *PRISM (Polarized Radiation Imaging and Spectroscopy Mission): an extended white paper*, *Journal of Cosmology and Astroparticle Physics* **2** (Feb., 2014) 6, [`arXiv:1310.1554`].

[78] J. Delabrouille, P. de Bernardis, F. R. Bouchet, et al., *Exploring cosmic origins with CORE: Survey requirements and mission design*, *Journal of Cosmology and Astroparticle Physics* **4** (Apr., 2018) 014, [`arXiv:1706.04516`].

[79] J. Delabrouille, M. H. Abitbol, N. Aghanim, et al., *Microwave spectro-polarimetry of matter and radiation across space and time*, *Experimental Astronomy* **51** (June, 2021) 1471–1514.

[80] J. Chluba and D. Grin, *CMB spectral distortions from small-scale isocurvature fluctuations*, *Monthly Notices of the Royal Astronomical Society* **434** (Sept., 2013) 1619–1635, [`arXiv:1304.4596`].

[81] T. R. Slatyer, *The Sommerfeld enhancement for dark matter with an excited state*, *Journal of Cosmology and Astroparticle Physics* **2010** (Feb., 2010) 028, [`arXiv:0910.5713`].

[82] A. Aguirre and M. C. Johnson, *A status report on the observability of cosmic bubble collisions*, *Reports on Progress in Physics* **74** (July, 2011) 074901, [`arXiv:0908.4105`].

[83] M. Kleban, T. S. Levi, and K. Sigurdson, *Observing the multiverse with cosmic wakes*, *Physical Review D* **87** (Feb., 2013) 041301, [`arXiv:1109.3473`].

[84] H. Deng, *Spiky CMB distortions from primordial bubbles*, *Journal of Cosmology and Astroparticle Physics* **2020** (May, 2020) 037, [`arXiv:2003.02485`].

[85] B. Carr, F. Kühnel, and M. Sandstad, *Primordial black holes as dark matter*, *Physical Review D* **94** (Oct., 2016) 083504, [`arXiv:1607.06077`].

[86] K. T. Abe, H. Tashiro, and T. Tanaka, *Thermal Sunyaev-Zel'dovich anisotropy due to primordial black holes*, *Physical Review D* **99** (May, 2019) 103519, [`arXiv:1901.06809`].

[87] H. Deng, *μ-distortion around stupendously large primordial black holes*, *Journal of Cosmology and Astroparticle Physics* **2021** (Nov., 2021) 054, [`arXiv:2106.09817`].

[88] O. Özsoy and G. Tasinato, *CMB μT cross correlations as a probe of primordial black hole scenarios*, *Physical Review D* **104** (Aug., 2021) 043526, [`arXiv:2104.12792`].

[89] K. Jedamzik, V. Katalinić, and A. V. Olinto, *Damping of cosmic magnetic fields*, *Physical Review D* **57** (Mar., 1998) 3264–3284, [`astro-ph/9606080`].

[90] K. Miyamoto, T. Sekiguchi, H. Tashiro, and S. Yokoyama, *CMB distortion anisotropies due to the decay of primordial magnetic fields*, *Physical Review D* **89** (Mar., 2014) 063508, [`arXiv:1310.3886`].

[91] T. Minoda, K. Hasegawa, H. Tashiro, K. Ichiki, and N. Sugiyama, *Thermal Sunyaev-Zel'dovich Effect in the IGM due to Primordial Magnetic Fields*, *Galaxies* **6** (Dec., 2018) 143, [`arXiv:1812.09813`].

[92] S. Saga, A. Ota, H. Tashiro, and S. Yokoyama, *Secondary CMB temperature anisotropies from magnetic reheating*, *Monthly Notices of the Royal Astronomical Society* **490** (Dec., 2019) 4419–4427, [`arXiv:1904.09121`].

[93] C. Dvorkin, K. Blum, and M. Kamionkowski, *Constraining dark matter-baryon scattering with linear cosmology*, *Physical Review D* **89** (Jan., 2014) 023519, [`arXiv:1311.2937`].

[94] Y. Ali-Haïmoud, J. Chluba, and M. Kamionkowski, *Constraints on dark matter interactions with standard model particles from CMB spectral distortions*, *ArXiv e-prints* (June, 2015) [`arXiv:1506.04745`].

[95] J. B. Muñoz, E. D. Kovetz, and Y. Ali-Haïmoud, *Heating of baryons due to scattering with dark matter during the dark ages*, *Physical Review D* **92** (Oct., 2015) 083528, [`arXiv:1509.00029`].

[96] T. R. Slatyer, N. Padmanabhan, and D. P. Finkbeiner, *Cmb constraints on wimp annihilation: Energy absorption during the recombination epoch*, *Physical Review D (Particles, Fields, Gravitation, and Cosmology)* **80** (2009), no. 4 043526.

[97] J. Chluba, *Could the cosmological recombination spectrum help us understand annihilating dark matter?*, *Monthly Notices of the Royal Astronomical Society* **402** (Feb., 2010) 1195–1207, [`arXiv:0910.3663`].

[98] E. Pajer and M. Zaldarriaga, *New Window on Primordial Non-Gaussianity*, *Physical Review Letters* **109** (July, 2012) 021302, [`arXiv:1201.5375`].

[99] E. Pajer and M. Zaldarriaga, *A hydrodynamical approach to CMB μ-distortion from primordial perturbations*, *Journal of Cosmology and Astroparticle Physics* **2** (Feb., 2013) 36, [`arXiv:1206.4479`].

[100] J. Chluba, E. Dimastrogiovanni, M. A. Amin, and M. Kamionkowski, *Evolution of CMB spectral distortion anisotropies and tests of primordial non-Gaussianity*, *Monthly Notices of the Royal Astronomical Society* **466** (Apr., 2017) 2390–2401, [`arXiv:1610.08711`].

[101] G. Cabass, E. Pajer, and D. van der Woude, *Spectral distortion anisotropies from single-field inflation*, *Journal of Cosmology and Astroparticle Physics* **2018** (Aug., 2018) 050, [`arXiv:1805.08775`].

[102] G. Orlando, P. D. Meerburg, and S. P. Patil, *Primordial tensor bispectra in μ-CMB cross-correlations*, *Journal of Cosmology and Astroparticle Physics* **2022** (Feb., 2022) 004, [`arXiv:2109.01095`].

[103] B. Audren, J. Lesgourgues, G. Mangano, P. D. Serpico, and T. Tram, *Strongest model-independent bound on the lifetime of Dark Matter*, *Journal of Cosmology and Astroparticle Physics* **2014** (Dec., 2014) 028–028, [`arXiv:1407.2418`].

A Perturbed decay term for decaying particle scenarios

For the average heating rate by particle decay we simply used

$$\frac{\mathrm{d}a^4\rho_\gamma^{(0)}}{a^4\,\mathrm{d}t} = -\frac{\mathrm{d}a^3\rho_X^{(0)}}{a^3\,\mathrm{d}t} = m_X c^2\,\Gamma_X N_X^{(0)}, \tag{A.1}$$

which follows from the Collision term for the decaying particle. Here, we assumed rapid transfer of energy to the photon field though Compton scattering. This then give the relevant heating term

$$\frac{\mathrm{d}Q^{(0)}}{\mathrm{d}t} = \frac{\mathrm{d}a^4\rho_\gamma^{(0)}}{a^4\rho_\gamma^{(0)}\,\mathrm{d}t} = \frac{m_X c^2\,\Gamma_X N_X^{(0)}}{\rho_\gamma^{(0)}} \tag{A.2}$$

for the background evolution with $N_X^{(0)} = N_X^{(0)}(t=0)\,\mathrm{e}^{-\Gamma_X t}$.

To obtain the perturbed decay term for decaying particle scenarios, we can similarly use the evolution equation given by [103]

$$\frac{\partial}{\partial\eta}\delta_X^{(1)} = -k\beta_X - 3\frac{\partial}{\partial\eta}\Phi^{(1)} - a\Gamma_X\,\Psi^{(1)} \tag{A.3}$$

for the perturbations in a decaying particle component, $\delta_X^{(1)} = N_X^{(1)}/N_X^{(0)}$. We switched to our convention for the sign of Φ, which is $-\Phi^{\mathrm{A}}$. Aside from the last term, this is simply like the standard dark matter equation [7]. The last term gives rise to extra energy release from changes in the local time. Together with contributions from perturbations in the density of the decaying particle, the release of energy to the photon field then is given by

$$\frac{\mathrm{d}a^4\rho_\gamma^{(1)}}{a^4\,\mathrm{d}t} = \delta_X^{(1)}\,\frac{\mathrm{d}a^4\rho_\gamma^{(0)}}{a^4\,\mathrm{d}t} + m_X c^2\,N_X^{(0)}\,\Gamma_X\,\Psi^{(1)} \equiv \left[\delta_X^{(1)} + \Psi^{(1)}\right]\frac{\mathrm{d}a^4\rho_\gamma^{(0)}}{a^4\,\mathrm{d}t}. \tag{A.4}$$

This implies

$$\frac{\mathrm{d}Q^{(1)}}{\mathrm{d}t} = \left[\delta_X^{(1)} + \Psi^{(1)}\right]\frac{\mathrm{d}Q^{(0)}}{\mathrm{d}t}, \tag{A.5}$$

which includes the particle density and local potential modulation effects. If we furthermore assume that $\delta_X \simeq \delta_{\mathrm{cdm}}$, we obtain the expression in Eq. (3.2). This expression can also be directly obtained when thinking of the corrections from $\Psi^{(1)}$ to the background collision term [35]. A more in depth derivation is given in the Appendix of paper II, which also includes effects from the local heat capacity modulation.

B Power spectrum convergence

Analogously to Fig. 3, we can study the convergence of the power spectra, and build confidence in the extended basis. Again we show the three redshifts corresponding to the three eras ($z = 5\times10^5$, $z = 5\times10^4$ and $z = 5\times10^3$), but with one additional residual-era injection which is typically the poorest converged in this formalism (i.e., $z = 10^5$, see paper I). We see in Fig. 32 and Fig. 33 that similar results hold as for studying individual spectra: the y-era injection requires only a single additional mode y_1 by construction, while the residual- and μ-era injections requires around 3-5 modes. Only the second panel shows an appreciable difference between $N_{\max} = 13$ and $N_{\max} = 15$, however this only amounts to a $\lesssim 2\%$ difference. These levels of departures are at the limit of our computations, on par with other neglected effects.

C Physical effects in SD power spectra

For completeness, we include the pure distortion power spectra with various physical *switches*, which helps illustrate the origin of concrete features. Many of these will be very analogous to those discussed in Sect. 4.3.1, but are useful to see in the absence of structure familiar from the temperature power spectra.

Referring to Figs. 34, 35 and 36, the first feature we note is the lack of thermalisation contributions for late injection. At early injection times thermalisation sources both $\mu^{(1)}$ and $y^{(1)}$ through the $M_{\rm D}\boldsymbol{y}_0^{(0)}$ and $(Y_1 - Y)$ terms. We independently verify perturbed photon emission only influences $\mu^{(1)}$, and is a small exclusively early time effect. Anisotropic heating exclusively sources $\mu^{(1)}$ at early times as expected, with other times sourcing a mix of both primary SEDs. We can also again see a characteristic ceasing of potential boosts at the late time, since much of the energy is injected either sub-horizon or very close to horizon crossing. Interestingly we see that the residual-era injection sees a remnant of the first power spectrum peak in both $\mu \times \mu$ *and* $y \times y$. For the later injection times that Doppler peak is lost in both cases, and at early times it is only present in the $y \times y$ spectrum given that boosting is a subdominant contribution to the $\mu^{(1)}$ SED in the presence of anisotropic heating and perturbed thermalisation.

D Correlations with E-modes

For completeness we show the correlations between distortion SEDs and temperature polarisation E-modes. While these are more difficult to interpret than the correlations with local temperature, they are a crucial contribution to the sensitivity within the forecasts. As seen in Fig. 37 these have complex and strongly time dependent patterns. By inspecting Fig. 38 this complexity can be attributed to a strong correlation driven through perturbed thermalisation and anisotropic heating which, unlike the cases seen earlier, are almost comparable in amplitude to the Doppler boosting term rather than being a small correction. Recalling that perturbed thermalisation and anisotropic heating largely cancel in the case of perturbed decay, this can explain the significant loss of power in the $\mu \times E$ (dashed blue lines) within Fig. 30 compared to Fig. 28.

As is very familiar by now, for late time injection these terms become less important and the signal relies more on the boosting mechanism. This together with the characteristic sign flip for $\mu^{(1)}$ for boosting early and late time injections yields a strong overall flip of the $\mu \times E$ power spectrum.

E Survey specifications

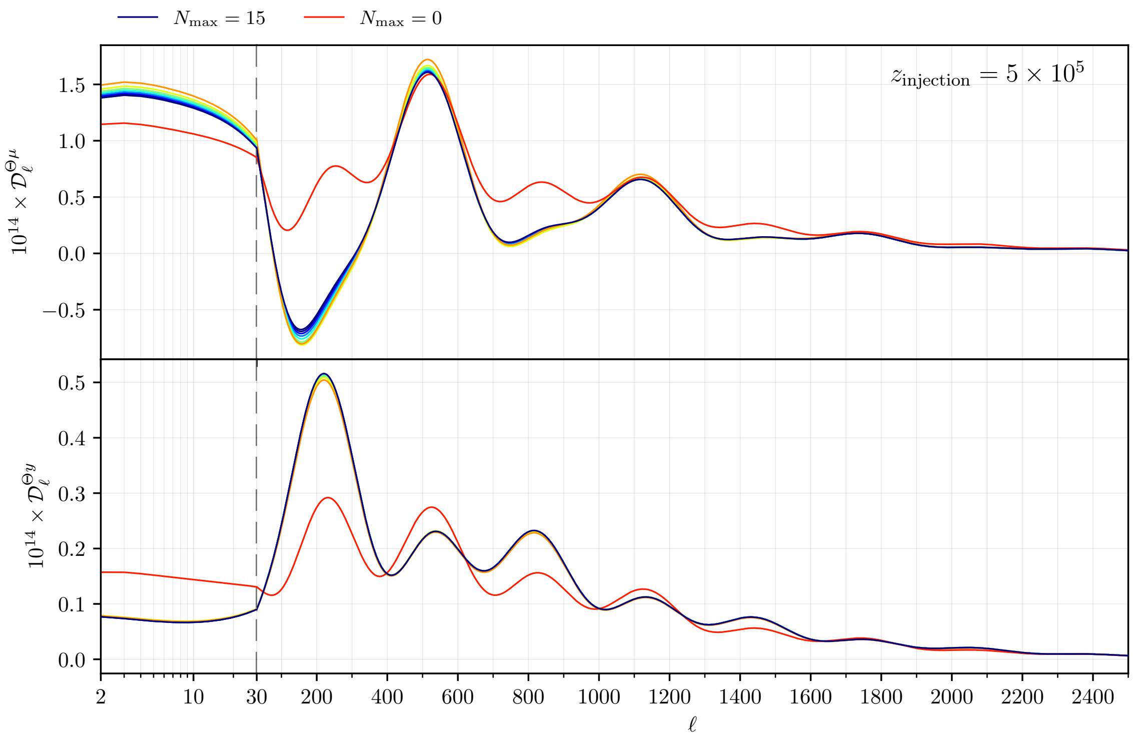
In table 1 we summarize the specifications of *Planck* that we use in the forecasts, in table 2 we do the same for *LiteBIRD* and in table 3 for *PICO*.

Channel [GHz]	Effective beam FWHM [arcmin]	Temperature Noise [μK arcmin]
30	32.29	150
44	27.94	162
70	13.08	210
100	9.66	77
143	7.22	33
217	4.90	47
353	4.92	154
545	4.67	818
857	4.22	30019

Table 1: *Planck* specifications from [59].

Channel [GHz]	Effective beam FWHM [arcmin]	Temperature Noise [μK arcmin]
40	70.5	$37.42 / \sqrt{2}$
50	58.5	$33.46 / \sqrt{2}$
60	51.1	$21.31 / \sqrt{2}$
68	41.6	$16.87 / \sqrt{2}$
78	36.9	$12.07 / \sqrt{2}$
89	33.0	$11.30 / \sqrt{2}$
100	30.2	$6.56 / \sqrt{2}$
119	26.3	$4.58 / \sqrt{2}$
140	23.7	$4.79 / \sqrt{2}$
166	28.9	$5.57 / \sqrt{2}$
195	28.0	$5.85 / \sqrt{2}$
235	24.7	$10.79 / \sqrt{2}$
280	22.5	$13.80 / \sqrt{2}$
337	20.9	$21.95 / \sqrt{2}$
402	17.9	$47.45 / \sqrt{2}$

Table 2: *LiteBIRD* specifications from [60].



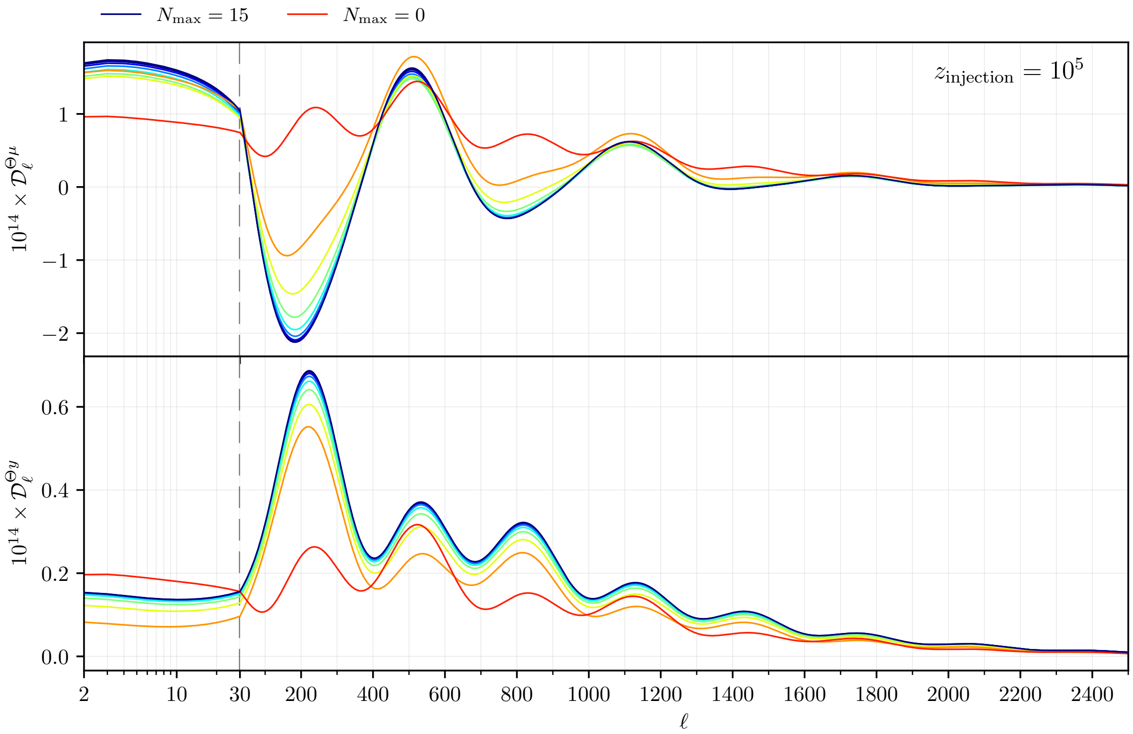


Figure 32: Two figures illustrating convergence of the power spectra for increasing $N_{\max}$. Shown are the $\Theta\times\mu$ and $\Theta\times y$ spectra at early injection redshifts. The vertical dashed line shows a division between linearly-spaced ℓ values (left) and log-spaced values (right).

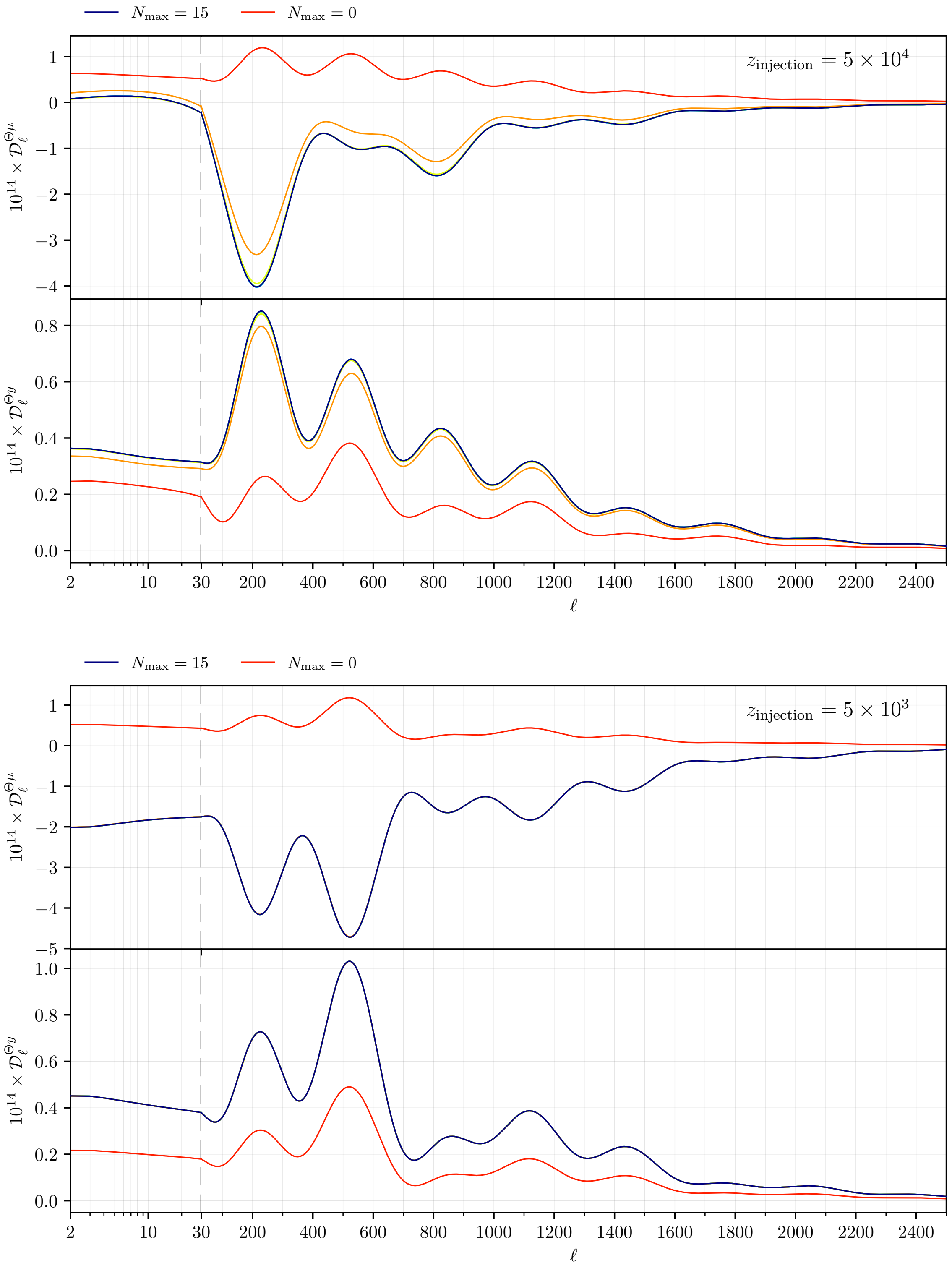


Figure 33: As for Fig. 32 but for late injection redshifts.

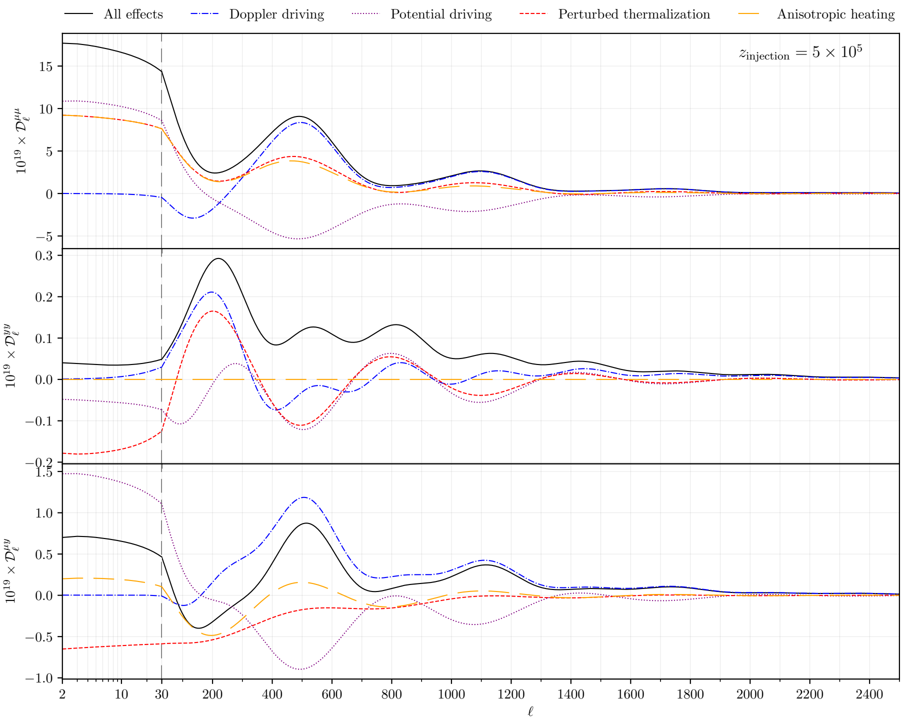


Figure 34: A figure showing the distortion power spectra from injection at $z_{\rm injection} = 5 \times 10^5$ with different physical switches.

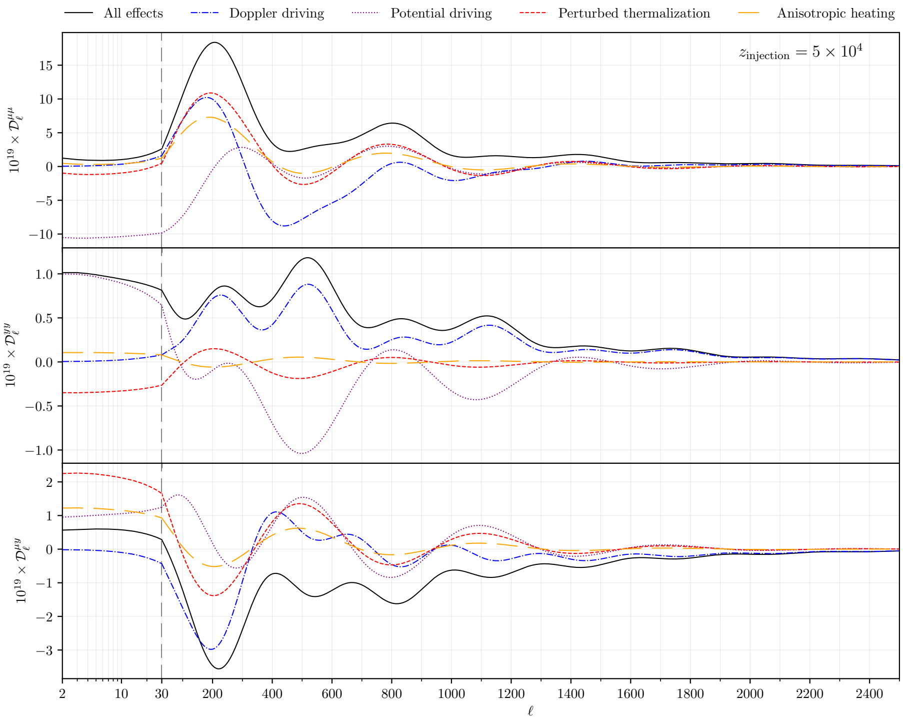


Figure 35: As for Fig. 34 but for $z_{\rm injection} = 5 \times 10^4$.

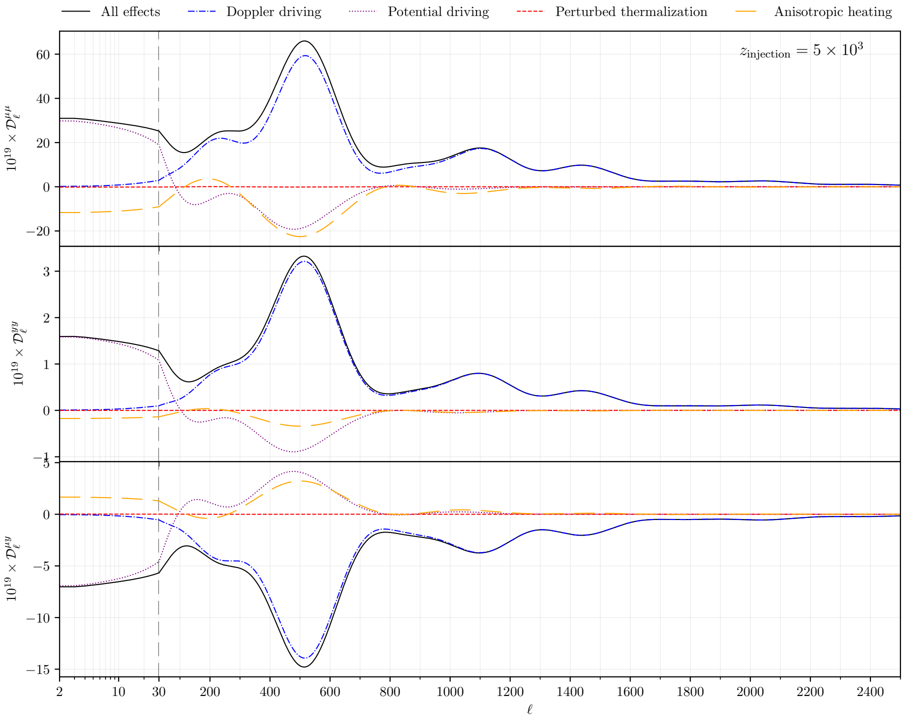


Figure 36: As for Fig. 34 but for $z_{\rm injection} = 5 \times 10^3$.

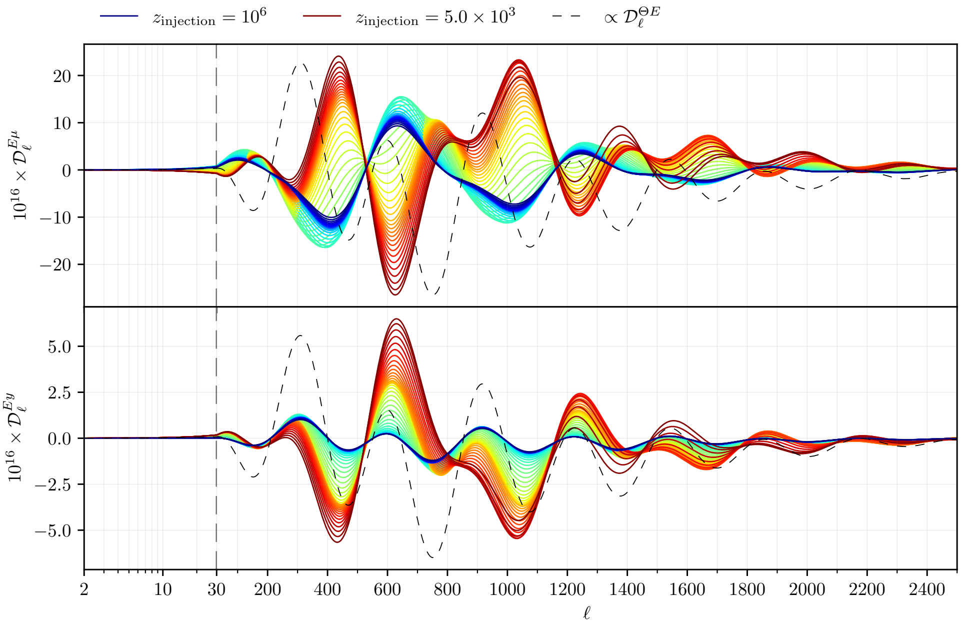


Figure 37: The power spectra for $E \times \mu$ and $E \times y$ over a range of 50 single-injection redshifts. Blue lines show early injection into the μ-era and red lines show late injection in the y-era. The vertical dashed line shows a division between log-spaced ℓ values (left) and linear-spaced values (right). For reference, we show the familiar $\Theta \times E$ power spectrum (rescaled within each panel).

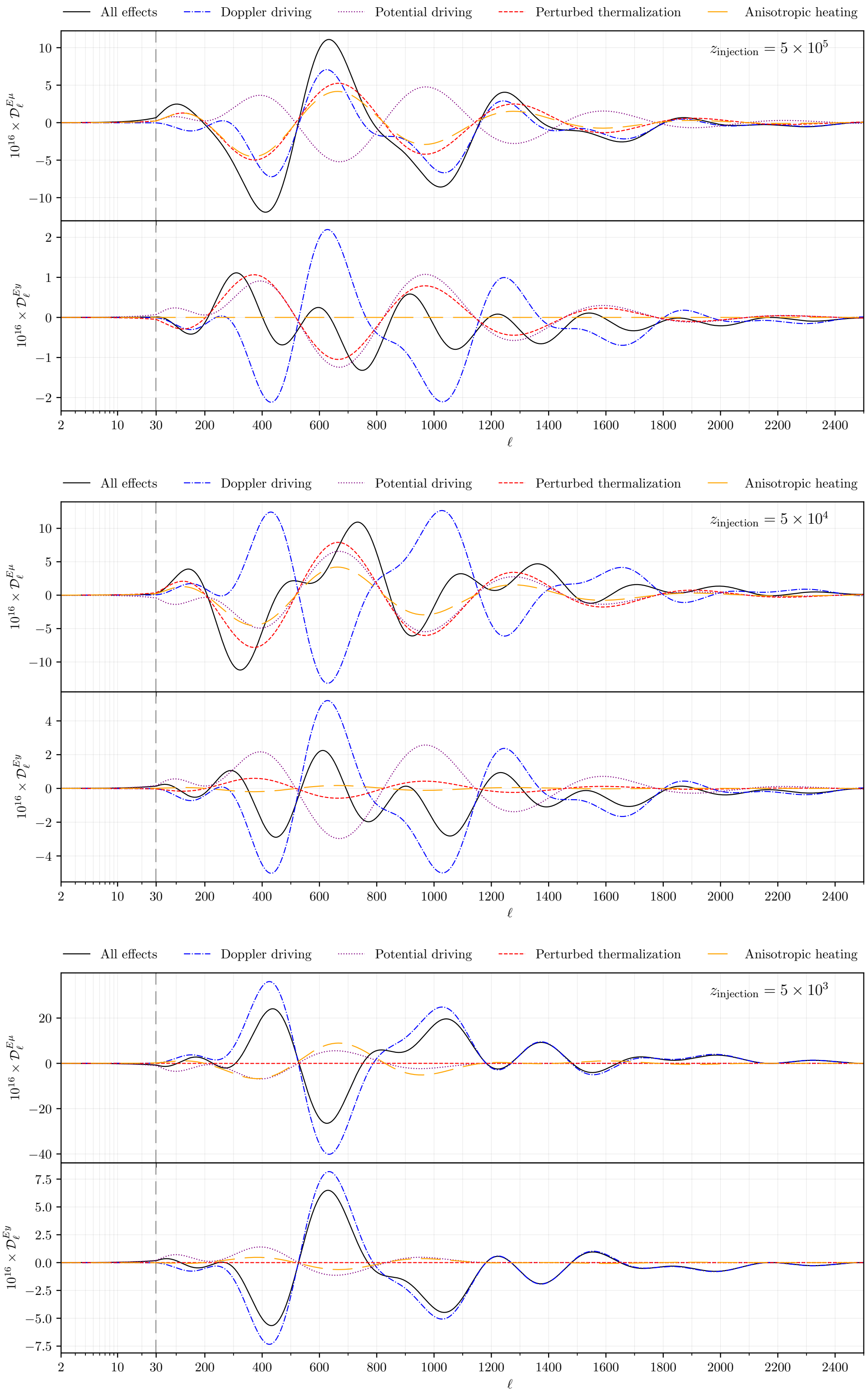


Figure 38: Three figures illustrating the $E \times \mu$ (top panel) and $E \times y$ (bottom panel) power spectra with various terms switched off. The figures from top to bottom show injection redshifts 5×10^5, 5×10^4 and 5×10^3. The vertical dashed line shows a division between linearly-spaced ℓ values (left) and log-spaced values (right).

Channel [GHz]	Effective beam FWHM [arcmin]	Temperature Noise [μK arcmin]
21	38.4	23.9 / $\sqrt{2}$
25	32.0	18.4 / $\sqrt{2}$
30	28.3	12.4 / $\sqrt{2}$
36	23.6	7.9 / $\sqrt{2}$
43	22.2	7.9 / $\sqrt{2}$
52	18.4	5.7 / $\sqrt{2}$
62	12.8	5.4 / $\sqrt{2}$
75	10.7	4.2 / $\sqrt{2}$
90	9.5	2.8 / $\sqrt{2}$
108	7.9	2.3 / $\sqrt{2}$
129	7.4	2.1 / $\sqrt{2}$
155	6.2	1.8 / $\sqrt{2}$
186	4.3	4.0 / $\sqrt{2}$
223	3.6	4.5 / $\sqrt{2}$
268	3.2	3.1 / $\sqrt{2}$
321	2.6	4.2 / $\sqrt{2}$
385	2.5	4.5 / $\sqrt{2}$
462	2.1	9.1 / $\sqrt{2}$
555	1.5	45.8 / $\sqrt{2}$
666	1.3	177.0 / $\sqrt{2}$
799	1.1	1050.0 / $\sqrt{2}$

Table 3: *PICO* specifications from [61].

9

Conclusion

In some sense the ultimate goal of this thesis was to understand Cosmological Perturbation Theory (CPT), which sits at the heart of modern early-universe Cosmology. It is this framework that allows us to understand the complex interactions within the plasma which filled the entire observable Universe for the first 380 thousand years. This framework has delivered extremely precise calculations of the statistics of temperature spots in the CMB sky, and the corresponding observational dataset has arguably pushed early-Universe research into the realm of a precision science. The bias towards early times present throughout this thesis should not detract from the importance ot CPT for the later Universe - the formalism studied here serves as the initial conditions for the perturbations which at later times collapse non-linearly and form the large scale structure we observe around us.

Understanding early universe perturbations first requires an understanding of the average Universe, both in its expansion (Sect. 1.2) and contents (Sect. 1.3). Both of these components receive corresponding inhomogeneities (Sect. 3.3 and Sect. 3.1 respectively) when studying perturbation theory, and they ultimately govern the anisotropies seen in the CMB sky.

We present an initially tangential topic in chapter 2, which involves a detailed look at the photon spectrum and relevant thermalisation processes which can create spectral distortions (SDs). We explain in that chapter how this impacts the average photon spectrum of the CMB sky, leaving the more novel and exciting implications for the following chapters.

With the introductory chapters in place, we can readily interpret the paper within chapter 4 as a study of tensorial perturbations and their sourcing of average an μ distortion in the CMB sky via a dissipation of energy into the primordial plasma. The following paper within chapter 5 is a follow up, less concerned with SDs, but instead focused on providing a demystifying close look at the tensor perturbations themselves.

The final three papers contained within this thesis (chapters 6-8) provide the most exciting and novel work, in essence uniting the discussions of chapters 2 and 3. The formalism introduced there extends the standard Boltzmann hierarchy to span the frequency domain, and thus allows for a full spectro-spatial view of the CMB sky.

While this thesis has concentrated on introducing a lot of theory, it is important to not lose sight of the consequences and impact of this research: the all important *why?* Recall the introduction to chapter 2, where it was emphasised that most of what we have from the early-Universe is a photon background. While sounding simple, this dataset is rich, and has allowed for tangible answers to some of the biggest questions regarding the structure and origin of the Universe. Spectral distortions allow for one way of studying this photon background even closer. The data from *COBE/FIRAS* allows us to constrain SD amplitudes to the order of 10^{-6}-10^{-5} (Fixsen et al., 1996), which in turn means we understand the thermodynamics of the primordial plasma to one part in $\sim$ 100 thousand ranging all the way back to 3 minutes after the big bang. That is an incredible statement in itself, and with modern technology we could do much better. One of the many reasons chapters 6-8 are so exciting is that they give us a new way of constraining these average SD amplitudes, and thus placing tighter limits on what processes could have happened in the primordial plasma.

The novelty of the spectro-spatial Boltzmann hierarchy simultaneously invites excitement for future exploration as well as humility in the results. While it is possible that the modelling presented could become a new standard for constraining early universe Physics (especially in the context of state of the art imagers pushing beyond the limits of *Planck*), it should also be recognised that much more work needs to be done to verify and consolidate this first push in a radically new direction.

One clear next step for building confidence in the new spectro-spatial Boltzmann hierarchy will be to study the equations analytically. This process will likely be much easier in certain limiting regimes, like for late time energy injection where SD patches are governed mostly by Doppler boosting of the average spectrum, or for low ℓ where most k modes were superhorizon till recent times. Furthermore, building an analytic understanding about the nature of energy injection is important, as this holds importance for both the gauge choice and adiabaticity of the perturbations. While analytic solutions would strongly increase the confidence in the results of chapter 8, it is worth noting that a holistic analytic understanding of standard CMB temperature anisotropies followed the numerical work by up to two decades[1].

Beyond developing analytic solutions a second numerical test can in principle be achieved with a more *brute force* approach. We have previously mentioned the numerical packages capable of calculating collision terms to high precision (Sarkar et al., 2019; Ravenni & Chluba, 2020; Chluba et al., 2020b), which within `CosmoTherm` can find spectral evolution for the average CMB (Chluba, 2013, 2015). In principle this binned frequency approach could be extended to modelling the anisotropic evolution. This is mentioned as a computationally expensive option in the introductory statements of chapter 6, and while it wouldn't be feasible to use this approach in future spectro-spatial CMB calculations it could provide some limited costly cross-checks.

Within the extended frequency hierarchy itself some clear improvements can be made. In chapter 6 we point out two ways in which the spectral basis is insufficient to replicate the full `CosmoTherm` results: firstly a delayed onset of the μ-era (excessive photon production), and secondly the basis not quite capturing the residual-era spectral shapes for energy injection around $z_{\text{injection}} \sim 10^5$. These problems motivate two improvements left to future work. Firstly revisiting the problem of photon production, which is known to have more sophisticated formulations which could improve the frequency hierarchy (Chluba, 2014a). Secondly studying better choices of spectral basis which capture vital information in the frequency space between the μ and y distortion peaks. This could be more physically and theoretically motivated (we considered briefly considered shapes following $\hat{O}_x^N(\mathcal{Y}_{\rm SZ}/x)$ which had promising mathematical properties) or simply a numerical choice of basis (e.g. built from PCA).

---

[1] Based on the introductory statements of (Hu & Sugiyama, 1995), where 5-10% agreement between analytics and numerics is achieved on all scales, while referencing numerical solutions from 25 years prior.

In brief then we can say that many directions remain to explore with this work. The papers presented in chapters 6-8 constitute the opening of a door to a new branch of work, not the closing of one.

> The surface of the Earth is the shore of the cosmic ocean. On this shore, we've learned most of what we know. Recently, we've waded a little way out, maybe ankle-deep, and the water seems inviting. Some part of our being knows this is where we came from. We long to return, and we can, because the cosmos is also within us. We're made of star stuff. We are a way for the cosmos to know itself.
>
> –Carl Sagan, *Cosmos*

Appendix A

Mathematical tools

A.1 Legendre transform

The legendre polynomials (Abramowitz & Stegun, 1965) $\mathcal{P}_\ell(x)$ make up a complete and orthogonal basis for functions on the interval $x \in [-1, 1]$. They are associated with angular distribution, with the interval usually intepreted as $x \in [\cos(\theta = 0), \cos(\theta = \pi)]$. With the basis being orthogonal we can use familiar mathematics to deduce the projection integrals. We start with an ansatz that[1]

$$f(x) = \sum_{\ell=0}^{\infty} \hat{f}_\ell \mathcal{P}_\ell(x), \tag{A.1}$$

and then assert that we can derive a given component $\hat{f}_{\ell_1}$ with the corresponding integral over $(2\ell + 1)\mathcal{P}_{\ell_1}/2$. This assertion holds due to the orthogonality condition

$$\int_{-1}^{+1} \mathrm{d}x \, \mathcal{P}_{\ell_1}(x) \mathcal{P}_{\ell_2}(x) = \frac{2}{2\ell + 1} \delta_{\ell_1 \ell_2}, \tag{A.2}$$

---

[1] We drop the Einstein summation convention, since here the indices have no relation to manifolds and curvature.

and can be seen as a generalisation of the notion that dotting a usual cartesian 3-vector with basis vector $\mathbf{e}_x$ gives the x component of the vector:

$$\hat{f}_{\ell_1} = \int_{-1}^{+1} \mathrm{d}x\, f(x)\, \frac{2\ell + 1)}{2} \mathcal{P}_{\ell_1}(x) \tag{A.3}$$

The bottom line here is that a function which depends on space can be decomposed into angle and length scale, with the former thus compressed into legendre moments $\Theta(x^\mu) \to \Theta(|x|)_\ell$. Additionally, any time this variable appears with an explicit factor of the angle (here x) we use the recursion relation

$$x\mathcal{P}_\ell(x) = \frac{1}{2\ell + 1} \left[(\ell + 1)\mathcal{P}_{\ell+1}(x) + \ell\mathcal{P}_{\ell-1}(x)\right]. \tag{A.4}$$

This appears, for example, in the collision term Eq. 3.4.

A.2 Spherical Harmonics

Similar to the Legendre polynomials we can use spherical harmonics $Y_{\ell m}(\theta, \phi)$ as a complete and orthogonal basis for angular space (Abramowitz & Stegun, 1965). In this case we have $X(\theta, \phi) = \sum_\ell \sum_{m=-\ell}^{\ell} X_{\ell m} Y_{\ell m}(\theta, \phi)$, where $X_{\ell m} = \int X(\theta, \phi) Y^*_{\ell m}(\theta, \phi)$. We won't make extensive use of these functions, however they appear implicitly in our formulation of cosmological perturbation theory (chapter 3) where we define $X_\ell = \sum_m X_{\ell m} Y_{\ell m}$.

A.3 Fourier transform

The fourier transform of function $f(x)$ is defined as

$$\tilde{f}(k) = \int f(x) e^{-ikx}\, \mathrm{d}x, \tag{A.5}$$

which can be understood as a projection of the function onto its corresponding plane wave representation exp(ikx). The inverse transform is similarly given by

$$f(x) = \frac{1}{2\pi} \int \tilde{f}(k) e^{ikx} \, \mathrm{d}k. \tag{A.6}$$

One important consequence of this transformation is that spatial derivatives are simply mapped to factors of the underlying wavenumber k. This simply follows from considering

$$\partial_x f(x) = \frac{1}{2\pi} \int \tilde{f}(k) \; \partial_x e^{ikx} \, \mathrm{d}k = ik \, \tilde{f}(k). \tag{A.7}$$

Outside of this appendix we will no longer explicitly label Fourier transforms with tildes.

References

Abazajian K. N. et al., 2015, Astroparticle Physics, 63, 66

Abbott B., et al., 2019, Phys. Rev. X, 9, 031040

Abramowitz M., Stegun I. A., 1965, Handbook of mathematical functions with formulas, graphs, and mathematical tables

Adler R. J., Casey B., Jacob O. C., 1995, American Journal of Physics, 63, 620

Arnett W. D., Bahcall J. N., Kirshner R. P., Woosley S. E., 1989, ARA&A, 27, 629

Bardeen J. M., 1980, Phys. Rev. D, 22, 1882

Birkhoff G. D., Langer R. E., 1923, Relativity and modern physics

Callin P., 2006, arXiv e-prints, astro

Caprini C., Figueroa D. G., 2018, Class. Quant. Grav., 35, 163001

Carroll S. M., 2004, Spacetime and geometry. An introduction to general relativity

Celoria M., Matarrese S., 2018, arXiv e-prints, arXiv:1812.08197

Chluba J., 2013, Monthly Notices of the Royal Astronomical Society, 434, 352

Chluba J., 2014a, Monthly Notices of the Royal Astronomical Society, 440, 2544

Chluba J., 2014b, arXiv e-prints, arXiv:1405.6938

Chluba J., 2015, Monthly Notices of the Royal Astronomical Society, 454, 4182

Chluba J., 2016, MNRAS, 460, 227

Chluba J., Dai L., Grin D., Amin M. A., Kamionkowski M., 2015a, Monthly Notices of the Royal Astronomical Society, 446, 2871

Chluba J., Grin D., 2013, MNRAS, 434, 1619

Chluba J., Hamann J., Patil S. P., 2015b, International Journal of Modern Physics D, 24, 1530023

Chluba J., Jeong D., 2014, Monthly Notices of the Royal Astronomical Society, 438, 2065

Chluba J., Khatri R., Sunyaev R. A., 2012, Monthly Notices of the Royal Astronomical Society, 425, 1129

Chluba J., Ravenni A., Acharya S. K., 2020a, arXiv e-prints, arXiv:2005.11325

Chluba J., Ravenni A., Bolliet B., 2020b, Monthly Notices of the Royal Astronomical Society, 492, 177

Chluba J., Sazonov S. Y., Sunyaev R. A., 2007, Astronomy and Astrophysics, 468, 785

Chluba J., Sunyaev R. A., 2004, Astronomy and Astrophysics, 424, 389

Chluba et al, 2019, Bulletin of the American Astronomical Society, 51, 184

Clifton T., Ferreira P. G., Padilla A., Skordis C., 2012, Phys. Rep., 513, 1

Dicus D. A., Repko W. W., 2005, Phys. Rev. D, 72, 088302

Dodelson S., 2003, Modern cosmology

Elwert G., Haug E., 1969, Physical Review, 183, 90

Fixsen D. J., Cheng E. S., Gales J. M., Mather J. C., Shafer R. A., Wright E. L., 1996, Astrophysical Journal, 473, 576

Gourgoulhon E., 2012, 3+1 Formalism in General Relativity, Vol. 846

Griffiths D. J., 2017, Introduction to Electrodynamics

Guth A. H., 1981, Phys. Rev. D, 23, 347

Hu W., Sugiyama N., 1995, ApJ, 444, 489

Hu W., Sugiyama N., 1996, ApJ, 471, 542

Hu W., White M., 1996, ApJ, 471, 30

Jauch J. M., Rohrlich F., 1976, The theory of photons and electrons. The relativistic quantum field theory of charged particles with spin one-half. Addison-Wesley Publishing Company

Kamionkowski M., 2021, Phys. Rev. D, 104, 063512

Kite T., Chluba J., Ravenni A., Patil S. P., 2022, MNRAS, 509, 1366

Kompaneets A. S., 1957, Soviet Journal of Experimental and Theoretical Physics, 4, 730

Lesgourgues J., 2011, arXiv e-prints, arXiv:1104.2932

Lewis A., Challinor A., Lasenby A., 2000, ApJ, 538, 473

LIGO Scientific Collaboration, Virgo Collaboration, 2016, Physical Review Letters, 116, 061102

Lucca M., Schöneberg N., Hooper D. C., Lesgourgues J., Chluba J., 2019, arXiv e-prints, arXiv:1910.04619

Ma C.-P., Bertschinger E., 1995, Astrophysical Journal, 455, 7

Mandl F., Skyrme T. H. R., 1952, Proceedings of the Royal Society of London Series A, 215, 497

Martínez V. J., Saar E., Martínez-González E., Pons-Bordería M. J., 2009, Data Analysis in Cosmology, Vol. 665

Misner C. W., Thorne K. S., Wheeler J. A., 2017, Gravitation

Mukhanov V. F., Feldman H. A., Brandenberger R. H., 1992, Phys. Rep., 215, 203

Ota A., Takahashi T., Tashiro H., Yamaguchi M., 2014, JCAP, 10, 29

Peebles P. J. E., Yu J. T., 1970, ApJ, 162, 815

Penzias A. A., Wilson R. W., 1965, The Astrophysical Journal, 142, 419

Pettinari G. W., 2016, The Intrinsic Bispectrum of the Cosmic Microwave Background

Planck Collaboration, 2018a, arXiv e-prints, arXiv:1807.06205

Planck Collaboration, 2018b, arXiv e-prints, arXiv:1807.06209

Planck Collaboration, 2019, arXiv e-prints, arXiv:1907.12875

Ravenni A., Chluba J., 2020, arXiv e-prints, arXiv:2005.06941

Remazeilles M., Chluba J., 2018, MNRAS, 478, 807

Rotti A., Chluba J., 2021, MNRAS, 500, 976

Saikawa K., Shirai S., 2018, JCAP, 2018, 035

Sarkar A., Chluba J., Lee E., 2019, Monthly Notices of the Royal Astronomical Society, 490, 3705

Seljak U., Zaldarriaga M., 1996, ApJ, 469, 437

Silk J., 1968, Astrophysical Journal, 151, 459

Stewart J. M., 1990, Classical and Quantum Gravity, 7, 1169

Tristram M. et al., 2022, Phys. Rev. D, 105, 083524

Wald R. M., 1984, General relativity

Watanabe Y., Komatsu E., 2006, Phys. Rev. D, 73, 123515

Weinberg S., 2004, Phys. Rev. D, 69, 023503

Weinberg S., 2006, Phys. Rev. D, 74, 063517

Weinberg S., 2008, Cosmology, Cosmology. OUP Oxford

Zeldovich Y. B., Sunyaev R. A., 1969, Astrophysics and Space Science, 4, 301