

Extracting the cosmic history from diffuse backgrounds

Thesis by

Jonathan R. Pritchard

Division of Physics, Mathematics, & Astronomy;

California Institute of Technology;

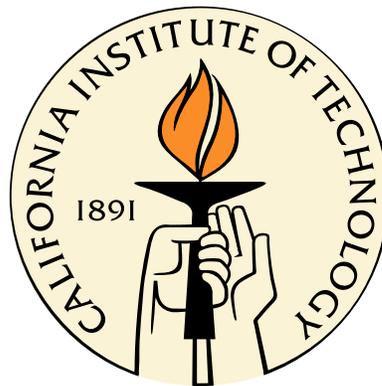
Mail Code 130-33;

Pasadena, CA 91125, USA

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy



California Institute of Technology

Pasadena, California

2007

(Submitted May 25, 2007)

© 2007

Jonathan R. Pritchard

Division of Physics, Mathematics, & Astronomy;

California Institute of Technology;

Mail Code 130-33;

Pasadena, CA 91125, USA

All Rights Reserved

True science teaches, above all, to doubt and be ignorant. –Miguel de Unamuno

Acknowledgements

Growing up in Somerset, near Stonehenge, its perhaps not surprising that the two recurring fascinations of my youth were with history and the heavens. On a clear night it was easy to lie shivering beneath the stars and wonder at what might be far away. Contrasting that sense of the far away was the concrete solidity of pre-history scattered about the landscape. Stonehenge, Avebury, and the many barrows and hill forts that spoke of the stone age and the depths of time. Being unable to be both an archeologist and an astronomer, my natural inclinations led me to cosmology. Here I get to look out and back at the same time, studying the archeology of the Universe itself. Somehow these simple beginnings have brought me to the completion of a doctorate at Caltech. No task requiring five years of work happens alone and I'd like to take this opportunity to thank those who have supported and encouraged my progress.

First, I'd especially like to thank my advisor, Marc Kamionkowski, for his advice, wisdom, and patience during the last five years here at Caltech. If not for his support during my SURF at Caltech I'd never have thought of applying for graduate school in the US. With his cryptic humour, he's perhaps not been the advisor I might have imagined, but almost certainly a much better one than I might have deserved. Thank you, Marc, for all that you've taught me about the world of science.

Second, I'd like to thank Steve Furlanetto, my long time collaborator here at Caltech. Working with him on reionization opened my eyes to an interesting period of cosmic history, which forms the bulk of this work.

I'd like to thank the members, past and present, of the TAPIR group, and especially everyone in Marc's cosmology group. They've been the people who provided a stimulating environment within which to do science. Thanks to JoAnn Boyd and Shirley Hampton for their administrative support. Thanks also to Chris Mach for keeping the computers running.

Last, but by no means least, I'd like to thank all of my friends here at Caltech for their friendship and support over the years: Lotty Ackerman, Dan Busby, Stephanie Chow, Cynthia Chiang, Mary Dunlop, Jon Eguia, Lisa Goggin, Katalin Grubits, Asa Hopkins, Asha Iyer, Katie Mack, Karin Menendez-Delmestre, Anthony Miller, Donal O'Connell, Francis O'Donovan, Paige Randall, Mike Salem, Sherry Suyu, Hareem Tariq, and Ramon van Handel.

To my long standing lunch partner and confidant Claire Jacobs: Thank you for keeping me sane.

To JB: Love and thanks for your interest in my exotic subject.

To my parents and sister: Your support while I've been on the other side of the world has been very important to me. Your pride reminds me that these abstract things I study are of interest to so many people in the world. Your love helps keep my feet upon the ground. To you all my thanks.

Abstract

The modern picture of the Universe resembles a detective novel with the first page and the middle chapters removed and the ending unwritten. Observations of the cosmic microwave background (CMB) have given cosmologists a snapshot of the Universe when it was only a few hundred thousand years old. At the same time, large galaxy surveys, such as SDSS and 2dF, have shed light on the distribution of matter in the local Universe. From the combination of these two data sets, cosmological parameters can be measured to percent accuracy. Two main frontiers remain: inflation, the domain of high-energy physics, and the epoch of reionization, the period connecting the linear age of the CMB with that of the present day. Added to this are the indications from supernovae of an acceleration in the expansion rate suggesting modifications to gravity or the presence of an esoteric new form of energy.

In this work, we investigate uses of various radiation backgrounds for probing the different epochs of this cosmic history. We examine (i) the use of B-mode polarization of the CMB induced by an inflationary gravitational wave background to probe inflation, (ii) the importance of higher Lyman series photons in pumping of the 21 cm line and the consequences for the 21 cm signal from the first stars, (iii) the atomic physics of Lyman series photon scattering in the intergalactic medium and the consequences for heating and coupling of the 21 cm line, (iv) the possibility of using the 21 cm line to probe inhomogeneous X-ray heating of the IGM by a population of early X-ray sources, and (v) the impact of inhomogeneous reionization on galaxy formation and the consequences for our ability to use large galaxy surveys to constrain dark energy. Together, these chapters significantly extend our understanding of important windows into the early Universe.

Contents

Acknowledgements	iv
Abstract	vi
1 Introduction	1
1.1 Our current knowledge of cosmic history	1
1.2 Cosmic microwave background fluctuations from gravitational waves: an analytic approach	2
1.3 Descending from on high: Lyman series cascades and spin-kinetic temperature coupling in the 21 cm line	4
1.4 The scattering of Lyman-series photons in the intergalactic medium	5
1.5 21 cm fluctuations from inhomogeneous X-ray heating before reionization	6
1.6 Galaxy surveys, inhomogeneous reionization, and dark energy	7
2 Cosmic microwave background fluctuations from gravitational waves: an analytic approach	9
2.1 Introduction	9
2.2 Exact equations	13
2.3 A tale of tensor modes	19
2.4 Gravitational-wave evolution	23
2.5 Recombination history	28
2.6 Projection factors	33
2.7 Source evolution	38
2.8 Discussion	44
2.9 Conclusions	46

3	Descending from on high: Lyman series cascades and spin-kinetic temperature coupling in the 21 cm line	47
3.1	Introduction	48
3.2	21 cm formalism and the Wouthuysen-Field mechanism	50
3.3	Direct pumping by Lyman series photons	54
3.4	Lyman series cascades	55
3.5	The Ly α coupling around a source	59
3.6	Brightness fluctuations from the first galaxies	62
3.6.1	Density fluctuations	62
3.6.2	Poisson fluctuations	66
3.6.3	Nonlinearities in the Wouthuysen-Field coupling	69
3.7	Conclusions	71
4	The scattering of Lyman-series photons in the intergalactic medium	73
4.1	Introduction	73
4.2	The radiation field near the Ly α resonance	76
4.3	The Ly α scattering rate	82
4.4	The colour temperature	86
4.5	Heat exchange from Ly α scattering	87
4.5.1	Continuous background	87
4.5.2	Injection at line center	89
4.6	Scattering of Ly n photons	91
4.7	Discussion	95
5	21 cm fluctuations from inhomogeneous X-ray heating before reionization	100
5.1	Introduction	100
5.2	21 cm signal	103
5.3	Global history	106
5.3.1	Outline	106
5.3.2	Heating and ionization	107
5.3.3	Ly α flux	112
5.3.4	Model histories	114
5.4	Formalism for temperature and ionization fluctuations	116

5.5	Temperature fluctuations	122
5.6	21 cm power spectrum	124
5.6.1	Redshift evolution	124
5.6.2	Spectral dependence	129
5.6.3	Effects of X-ray background	132
5.7	Observational Prospects	133
5.8	Conclusions	136
6	Galaxy surveys, inhomogeneous reionization, and dark energy	139
6.1	Introduction	139
6.2	Bubble model	142
6.3	Galaxy power spectrum	147
6.4	Fisher matrix	149
6.5	Possibility of detecting bubbles	153
6.6	Implications for dark-energy constraints	157
6.7	Conclusions	161
A	Numerical evolution of gravitational wave amplitude	173
A.1	Numerical evolution of gravitational waves with anisotropic stress	173
A.2	WKB solution	176
B	Calculating Einstein A coefficients in the hydrogen atom	182

List of Figures

1.1	CMB Polarization	4
1.2	Cartoon of different phases of 21 cm signal in redshift	5
2.1	Decomposition of temperature inisotropy into ISW and source terms	20
2.2	Evolution of the gravitational-wave amplitude	22
2.3	Tensor power spectra: TT, EE, and BB	23
2.4	Tensor TT and BB power spectra for approximations of the gravitational wave amplitude	29
2.5	Visibility function and optical depth during recombination	31
2.6	Gaussian approximation to the visibility function	32
2.7	Projection terms	34
2.8	Contribution to the power spectra due to the source term $g\Psi$	35
2.9	Comparison of power spectra with exact and approximate projection factors.	37
2.10	Comparison of approximations for the source term Ψ	40
2.11	Effect of phase-damping on the integrated source term	41
2.12	Comparison of EE and BB power spectra for exact and approximate calculations	43
3.1	Hyperfine structure of the $2P$ and $1S$ level of the hydrogen atom	52
3.2	Energy level diagram for the hydrogen atom illustrating $\text{Ly}\beta$ and $\text{Ly}\gamma$ cascades	56
3.3	Recycling fractions for $\text{Ly}n$ photons	58
3.4	Flux profile of a galaxy of mass $M = 3 \times 10^{10} M_{\odot}$ at $z = 20$ as a function of comoving distance	61
3.5	21 cm fluctuation power spectra and transfer function $W(k)$	64
3.6	Ratio of $W(k)$ calculated using proper atomic physics to $W(k)$ with $f_{\text{recycle}} = 1$	65
3.7	Correlation function for the Poisson fluctuations	68
3.8	Power spectrum for the Poisson fluctuations	68

4.1	Background radiation field near the Ly α resonance at $z = 10$	79
4.2	Ratio of δ_J in the “wing” approximation to the exact results	81
4.3	Scattering integral as a function of IGM temperature	84
4.4	Scattering integral as a function of $\gamma = \tau_{\text{GP}}^{-1}$	85
4.5	Heating integral for continuous injection	85
4.6	Heating integral for continuous injection	89
4.7	As Fig. 4.6, but for injection at line center	90
4.8	As Fig. 4.5, but for injection at line center	91
5.1	Mean IGM thermal history	114
5.2	Ionization histories	115
5.3	Transfer function for Ly α and T_K fluctuations	120
5.4	Fluctuation power spectra for T_K	122
5.5	Temporal evolution of the fluctuations in T_K	123
5.6	Temporal evolution of brightness temperature fluctuations	125
5.7	Full T_b power spectra for Model A in range $z = 17 - 20$	126
5.8	Full T_b power spectra for Model A in range $z = 13 - 16$	128
5.9	Effect of X-ray spectra on 21 cm power spectra	130
5.10	Comparison of Ly α induced power spectra for different Ly α sources	131
5.11	Effect of X-ray luminosity on evolution of $ \bar{T}_b \bar{\Delta}_{T_b}(k)$	133
5.12	Predicted error bars on T_b power spectrum for an SKA like instrument	134
6.1	Comparison of the galaxy and bubble power spectra	144
6.2	Model A: Contour map of detection	154
6.3	Model B: Contour map of detection	154
6.4	Model C: Contour map of detection	155
6.5	Illustration of the reduced covariance matrix	158
6.6	Model A: Contour map of errors in w_0 in the bubble parameter plane	160
6.7	Model B: Contour map of errors in w_0 in the bubble parameter plane	160
6.8	Model C: Contour map of errors in w_0 in the bubble parameter plane	161
A.1	Transfer function for gravitational waves when anisotropic stress is present	176
A.2	T and B power spectra with and without anisotropic stress	177

A.3 Evolution of gravitational wave amplitude for WKB solution 181

List of Tables

3.1	Recycling fractions and decay probabilities to the ground state	57
4.1	The quantity $(1 - S_\alpha)$ as a function of temperature T_K at $z = 20$ under several approximations	96
4.2	The quantity I_c as a function of temperature T_K at $z = 20$ under several approximations	97
6.1	Specification for CMB experiments	152
6.2	Specification for galaxy surveys	152

Chapter 1

Introduction

1.1 Our current knowledge of cosmic history

One of the goals of cosmology is to develop a narrative describing the complete history of the Universe. The advent of new observational techniques and technology within the last decade has turned hazy creation myth into detailed story. Yet like an unfinished manuscript many sections of the story remain unclear. Happily, unlike archaeology, cosmology benefits from having a device for peering into the past. Exploiting cosmological redshifting, observations at different frequencies may be used to puzzle out the full story of the Universe.

The standard picture that has emerged owes much to precise satellite observations of the cosmic microwave background (CMB) (Mather et al., 1990; Spergel et al., 2006). The uniformity seen in this snapshot of the Universe, at about four hundred thousand years after the big bang, provides a motivation for inflation (Guth, 1981). This period of rapid expansion lasts for only a fraction of a second, but provides the initial conditions for future evolution. As yet, we have only indirect evidence for inflation and little understanding of the underlying physical mechanism. The small CMB temperature anisotropies observed by WMAP (Spergel et al., 2006) have provided a powerful probe, not only of the physics of recombination, but of the global nature of the Universe.

The slight density inhomogeneities implied by CMB observations provide the initial seeds of structure that eventually grow into the cosmic web of galaxies seen today. The gravitational instability amplifies initially small overdensities until eventually these regions break off from the Hubble expansion to form gravitationally bound haloes, which host galaxies. Large galaxy surveys, such as 2dF and SDSS, provide a picture of luminous structure today, which is in agreement with the CMB. Comparing the two indicates that

much of the energy density of the Universe is in the form of a cosmological constant or dark energy that drives an acceleration of the Universe’s expansion.

In combination, these two large data sets—CMB and large scale structure (LSS)—provide a broad brush picture of the Universe’s evolution. Yet they are separated by three orders of magnitude in redshift and much physics is expected to occur in between. At around 150 million years after the big bang, the first stars form from the collapse of cold dense fragments of gas (Barkana & Loeb, 2001). As more stars and, eventually, galaxies form, their radiation affects gas in the intergalactic medium (IGM). This radiation can heat and ionize that gas affecting future galaxy formation. Eventually, expanding bubbles of ionized gas, surrounding luminous sources of ionizing radiation, overlap leading to a phase transition—reionization—from a neutral to an ionized IGM. Although reionization must occur, the details of its spatial and temporal structure are currently unknown (Barkana & Loeb, 2001). Similarly, the nature of the first stars—their structure and where they form—is poorly understood. To answer these questions, new observational techniques are required capable of probing the period from redshift $z = 6 - 1000$.

The rest of this chapter summarises the contents of this thesis, which consists of five previously published papers (Pritchard & Kamionkowski, 2005; Pritchard & Furlanetto, 2006; Furlanetto & Pritchard, 2006; Pritchard & Furlanetto, 2007; Pritchard et al., 2006) reproduced here with permission and arranged in rough redshift order. In Chapter 2, we summarise our theoretical knowledge of detecting inflationary gravitational waves via polarization of the CMB. Chapters 3, 4, and 5 explore the physics of the 21 cm line of neutral hydrogen and possibilities for using this to probe radiation from the first generations of stars. Finally, in Chapter 6, we look at the effect of inhomogeneous reionization on galaxy formation and consider the implications for constraining dark energy. Two appendices, taken from the papers, contain further information about calculating the evolution of tensor modes and Einstein A coefficients.

1.2 Cosmic microwave background fluctuations from gravitational waves: an analytic approach

The inflationary paradigm has been invoked to explain the observed flatness of the Universe and the origin of density perturbations (Guth, 1981). Alongside the production of density

fluctuations, many models of inflation also predict the production of gravitational waves (Abbott & Wise, 1984; Rubakov et al., 1982; Fabbri & Pollock, 1983; Starobinsky, 1985). The detection of these inflationary tensor modes would provide vital confirmation of the inflationary paradigm and could possibly shed light on particle physics at the highest energy scales. Detecting these gravitational waves directly will likely require the next generation successors to LIGO and LISA (Smith et al., 2006). In the nearer term there exists the possibility of indirect detection of inflationary gravitational waves via polarization of the CMB. A key feature of the polarization field is that it can be decomposed into E and B modes (see Figure 1.1), of which B modes are produced only by gravitational waves (Kamionkowski et al., 1997; Zaldarriaga & Seljak, 1997).

In Chapter 2, we examine the physical processes that lead to the production of a B-mode signal in the CMB (Kamionkowski et al., 1997; Zaldarriaga & Seljak, 1997). From this analysis, we develop analytic approximations that aid intuitive understanding of the B-mode power spectrum. We also develop a set of scaling relationships for the power spectrum in different regimes. The starting point for this is to consider the evolution of tensor modes through a first radiation-, then matter-dominated Universe. It is straightforward to find analytic solutions in the two extremes, but also useful to use a WKB approach to connect them. In addition, we consider the effect of anisotropic stress from free-streaming neutrinos (Weinberg, 2004). A numeric calculation shows that these damp those gravitational waves that enter the horizon during radiation domination.

Next, we develop the perturbation theory in the tight coupling approximation to connect polarization multipoles to the gravitational wave amplitude. We show that the tight coupling approximation needs to be supplemented with an exponential damping term introduced by the finite width of the surface of last scattering. This phase damping washes out the signal on small scales. Incorporating this physics allows us to develop an analytic form for the polarization multipoles at the surface of last scattering. To connect these multipoles to the anisotropies seen in the CMB today, we must project functions of wavenumber onto angular scales. These projection factors oscillate rapidly compared to the source function and we show that useful approximations may be found by applying a result of Debye (1909) and averaging over the oscillations. Putting all of these approximations together allows the B-mode power spectrum to be calculated analytically with reasonable precision, indicating that we have incorporated all of the essential physics.

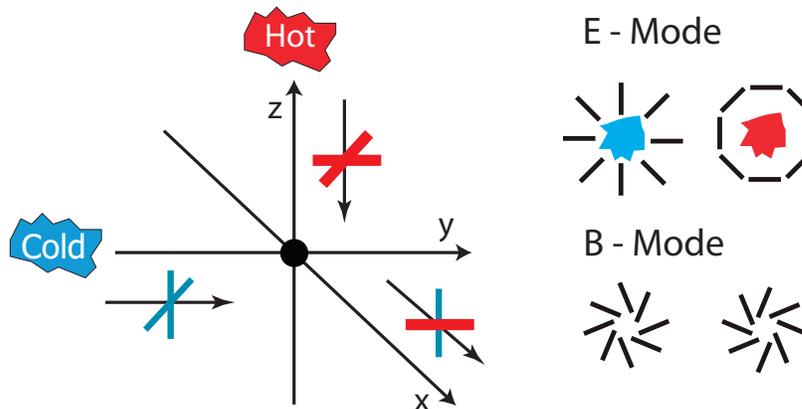


Figure 1.1: CMB Polarization. *Left:* Thompson scattering of an anisotropic radiation field generates polarization. *Right:* Polarization patterns can be decomposed into curl-free (E-mode) and curl (B-mode) patterns.

1.3 Descending from on high: Lyman series cascades and spin-kinetic temperature coupling in the 21 cm line

Observations of the 21 cm hyperfine transition of neutral hydrogen may open a new window onto the properties of the first stars and the beginning of the period of “cosmic twilight” (see Figure 1.2). Understanding the results of these observations will require a detailed understanding of the atomic physics involved in generating the 21 cm signal. As $\text{Ly}\alpha$ pumping (Wouthuysen, 1952; Field, 1959a) will be the dominant mechanism coupling gas and spin temperatures, it is vital that all aspects of $\text{Ly}\alpha$ photon production and coupling be understood.

In Chapter 3, we discuss the contribution of higher Lyman series photons to coupling of the spin temperature to the gas temperature. There are two ways $\text{Ly}n$ photons might contribute: directly, in a manner analogous to the Wouthuysen-Field effect, or indirectly, by producing $\text{Ly}\alpha$ photons as the result of an atomic cascade. We show that the direct contribution is negligible, as $\text{Ly}n$ photons scatter only a few times before undergoing an atomic cascade, while $\text{Ly}\alpha$ photons may scatter several million times. We then proceed to explore atomic cascades in more detail.

When a $\text{Ly}n$ photon is absorbed, the excited state produced may relax directly to the ground state or to another excited state. In the latter case an atomic cascade will result, ending in either the production of a $\text{Ly}\alpha$ photon or a two-photon decay from the 2S level. We

show that around 30% of $\text{Ly}\eta$ photons are converted into $\text{Ly}\alpha$ photons by atomic cascades. As a typical star will produce approximately the same number of photons between the $\text{Ly}\alpha$ and $\text{Ly}\beta$ resonances as between $\text{Ly}\beta$ and the Lyman limit, the $\text{Ly}\alpha$ flux generated by $\text{Ly}\eta$ photons can be significant.

Having explored the atomic physics, we place it in the context of the very first generation of sources. We convert the $\text{Ly}\alpha$ profile about a source into a power spectrum of 21 cm fluctuations using the formalism of Barkana & Loeb (2005b). This serves to illustrate the importance of including the correct atomic physics showing that $\text{Ly}\eta$ photons make a considerable correction to just including $\text{Ly}\alpha$ photons.

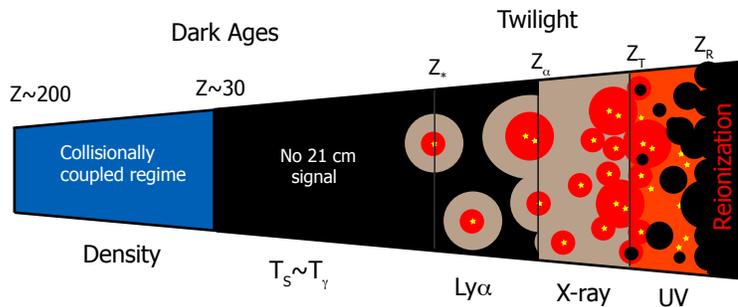


Figure 1.2: Schematic break down of contributions to 21 cm signal with redshift.

1.4 The scattering of Lyman-series photons in the intergalactic medium

When Lyman series photons scatter from hydrogen in the intergalactic medium they can heat the gas. Additionally, they can provide a mechanism for coupling the 21 cm spin temperature to the gas temperature. The relative magnitude of these effects has significant consequences for the observed 21 cm signal. It was initially thought (Madau et al., 1997) that scattering of $\text{Ly}\alpha$ photons would heat the gas above the CMB temperature before coupling became important. This would rule out the possibility of seeing the 21 cm signal in absorption. However, Chen & Miralda-Escudé (2004) showed that scattering of $\text{Ly}\alpha$ photons altered the distribution of photon frequencies near the $\text{Ly}\alpha$ resonance, greatly reducing the heating rate. Ensuring a correct calculation of the physics of scattering $\text{Ly}\alpha$ photons is vital for predicting the 21 cm signal. We argued in Pritchard & Furlanetto

(2006) that higher Lyman series photons may be important for coupling, but are probably not important for heating the gas.

In Chapter 4, we consider this question in more detail. We build upon previous work (Rybicki, 2006; Chuzhoy & Shapiro, 2006b; Meiksin, 2006) to develop analytic approximations for the contribution to heating and coupling by $\text{Ly}\alpha$ and $\text{Ly}\eta$ photons. We show that, while on a photon-by-photon basis $\text{Ly}\eta$ photons are more effective at transferring heat to the IGM, because they scatter only a few times before undergoing an atomic cascade, $\text{Ly}\eta$ photons do not heat the IGM significantly. A comparison between our analytic formulae and numerical calculation illustrates the accuracy of the “wing” approximation.

1.5 21 cm fluctuations from inhomogeneous X-ray heating before reionization

As the first stars are born and die they leave behind stellar remnants that may serve as X-ray sources; e.g., mini-quasars, SN remnants, and X-ray binaries. The resulting X-ray background will likely provide the dominant source of heating in the IGM before reionization. Many authors have taken the view that, because hard X-rays have a long mean free path (comparable to the Hubble size), this heating will be uniform. In fact, while there may be a uniform component, most of the energy from these X-ray sources is emitted as soft X-rays, which will be absorbed over Mpc scales (Furlanetto et al., 2006). As such, the heating will be very inhomogeneous.

In Chapter 5, we consider the consequences of inhomogeneous X-ray heating on the IGM, with a view to calculating the 21 cm signal from temperature fluctuations. These fluctuations may dominate the 21 cm signal before reionization gets underway (see Figure 1.2). We extend the formalism of Barkana & Loeb (2005b) to calculate fluctuations in the X-ray flux, which connect simply to fluctuations in the heating rate. To get temperature fluctuations, we must account for the interaction between adiabatic cooling and inhomogeneous heating. By numerically evolving temperature fluctuations from recombination, we show that as X-ray heating becomes significant temperature fluctuations track fluctuations in the heating rate. This calculation shows that there is a significant difference between inhomogeneous heating, which drives temperature fluctuations, and uniform heating, which tends to wash them out. This should be readily apparent in observations of the power

spectrum of gas temperature.

Having provided a strong motivation for the importance of distinguishing these two cases, we calculate the resulting 21 cm signal and show that it has a number of important properties. Most important is the existence of a trough in the absorption signal, which arises from competition between density and temperature fluctuations. The presence or absence of this trough helps determine whether the gas is cooler or hotter than the CMB, potentially allowing constraints on the gas thermal history.

The 21 cm signal contains many contributions from fluctuations in density, Ly α flux, gas temperature, and neutral fraction. If all these quantities fluctuate simultaneously then separating them out to obtain detailed astrophysical information will be very difficult. We use our machinery to calculate the evolution of the 21 cm power spectrum when both Ly α flux and gas temperature fluctuations contribute on top of the density field. This allows us to show that it is realistic to assume that the different contributions to the 21 cm signal may separate out with redshift, but that this separation is very sensitive to the astrophysical parameters. We consider the possibility of using the Square Kilometer Array to follow this evolution and conclude that provided foreground removal can be achieved the forecast is optimistic.

1.6 Galaxy surveys, inhomogeneous reionization, and dark energy

In Chapter 6, I turn to the nearby Universe. Large-scale galaxy surveys such as 2dF and SDSS allow us to infer the distribution of matter in the Universe. This works because it is possible to connect the underlying dark matter distribution to the distribution of galaxies. It is said that galaxies are a biased tracer of the density field. Provided that the bias is independent of scale then the shape of the galaxy and dark matter power spectra will be identical. We explore one possible mechanism that would produce a scale-dependent bias by looking at the effect of inhomogeneous reionization on galaxy formation.

During reionization, clusters of sources generate large (Mpc scale) HII regions, which are hotter than the average (Furlanetto et al., 2004). It is reasonable to assume that this environmental difference may impact the ability of gas to clump, cool, and form galaxies. In this chapter we specify a toy model linking the observed number of galaxies to the ionization

fraction during reionization. Thus bubbles imprint their pattern on the distribution of galaxies. This is somewhat hypothetical as the galaxies seen by surveys are formed by the merger of the many small galaxies that would form during reionization. Connecting the different generations of galaxies is complicated and we do not try, aiming instead to explore the consequences of an imprint and to determine whether it might be important.

If such an imprint exists then it might allow an independent probe of reionization or alternatively complicate attempts to use galaxy surveys to obtain cosmological parameters. Specifically, we explore the effect of the imprint on constraining dark energy parameters. Exploiting a Fisher matrix methodology to quantify the ability of galaxy surveys to distinguish between cosmology and a reionization imprint, we show that the imprint will probably not be important for current surveys. If the imprint is large enough to significantly bias determination of cosmological parameters then it is large enough to be seen and removed.

It may prove possible to use future galaxy surveys to probe reionization using such an imprint. For this purpose galaxy surveys at high redshift provide a good lever arm, as density structure has had less time to grow. If the imprint has a characteristic scale of tens of megaparsec, not impossible given current models of reionization, then a large survey at $z \gtrsim 3$ might be able to detect a signal.

Chapter 2

Cosmic microwave background fluctuations from gravitational waves: an analytic approach

We develop an analytic approach to calculation of the temperature and polarisation power spectra of the cosmic microwave background due to inflationary gravitational waves. This approach complements the more precise numerical results by providing insight into the physical origins of the features in the power spectra. We explore the use of analytic approximations for the gravitational-wave evolution, making use of the WKB approach to handle the radiation-matter transition. In the process, we describe scaling relations for the temperature and polarisation power spectra. We illustrate the dependence of the amplitude, shape, and peak locations on the details of recombination, the gravitational-wave power spectrum, and the cosmological parameters, and explain the origin of the peak locations in the temperature and polarisation power spectra. The decline in power on small scales in the polarisation power spectra is discussed in terms of phase-damping. In an appendix we detail numerical techniques for integrating the gravitational-wave evolution in the presence of anisotropic stress from free-streaming neutrinos.

Originally published as Pritchard and Kamionkowski, *Ann. Phys.*, **318**, 2 (2005).

2.1 Introduction

The standard hot big-bang cosmological model developed in the mid-twentieth century explained the expansion, the light-element abundances, and the cosmic microwave background (CMB) observed at a temperature $T = 2.7$ K. However, this model still left a number of questions unanswered. For example, the horizon problem: Why did the $\sim 4 \times 10^4$ causally disconnected regions probed by the CMB have the same temperature to one part in 10^5 ?

Where did the primordial seeds for large-scale structure (galaxies, clusters of galaxies, etc.) come from? And why did the Universe appear to be so flat?

In the early 1980s, inflation (Guth, 1981; Linde, 1982a; Albrecht & Steinhardt, 1982; Guth & Pi, 1982; Hawking, 1982; Linde, 1982b; Starobinsky, 1982; Bardeen et al., 1983; Lyth & Riotto, 1999), a paradigm in which an accelerated expansion is driven well within the first second after the big bang, was proposed as a solution to all of these problems. A relatively simple toy model, in which the expansion is driven by the vacuum energy associated with the displacement of some scalar field (the “inflaton”) from the minimum of its potential $V(\phi)$, could solve the horizon problem while simultaneously producing a flat Universe with a nearly scale-invariant spectrum of primordial density perturbations, a spectrum close to that required to explain the origin of galaxies, clusters of galaxies, and structure on even larger scales.

Over the past half decade, a suite of experiments (de Bernardis et al., 2000; Miller et al., 1999; Hanany et al., 2000; Halverson et al., 2002; Mason et al., 2003; Benoît et al., 2003; Goldstein et al., 2003; Spergel et al., 2003) have now measured the intensity of the CMB as a function of position on the sky with a resolution of a fraction of a degree over the entire sky and down to a few arcminutes over smaller patches. The power spectrum of the temperature fluctuations uncovered by these experiments has thus been measured precisely. Given a primordial spectrum of density perturbations like those predicted by inflation, it is straightforward to calculate the expected power spectrum, and this power spectrum exhibits a series of wiggles (“acoustic oscillations”), as a function of (angular) wavenumber or multipole moment l (Kamionkowski & Kosowsky, 1999). These oscillations have been well-studied theoretically, and the detailed features (e.g., heights, locations) can be used to determine the geometry of the Universe (Kamionkowski et al., 1994; Jungman et al., 1996a) and the values of cosmological parameters (Jungman et al., 1996b), as well as the spectrum of primordial perturbations. The experimental results are now conclusive: Structure formed from a nearly scale-invariant spectrum of primordial perturbations, and the Universe is constrained to be very close to, if not precisely, flat (de Bernardis et al., 2000; Miller et al., 1999; Hanany et al., 2000; Halverson et al., 2002; Mason et al., 2003; Benoît et al., 2003; Goldstein et al., 2003; Spergel et al., 2003; Kamionkowski et al., 1994; Efstathiou et al., 2002) .

Now that inflation has passed these tests, the obvious next step is to test the predic-

tions of inflation with higher precision, look for new tests of inflation, and perhaps try to determine the new physics of inflation, or in other words, the identity of the inflaton. In addition to more precise measurements of the power spectrum, there are also measurements of non-Gaussian correlations in the primordial density field (Bartolo et al., 2004) which, although expected to be small, should be nonzero. Another target for experiment is the inflationary-gravitational-wave (IGW) background. In addition to predicting primordial density perturbations and a flat Universe, inflation also predicts the existence of a nearly scale-invariant stochastic background of gravitational waves (Abbott & Wise, 1984; Rubakov et al., 1982; Fabbri & Pollock, 1983; Starobinsky, 1985), produced by quantum excitation of gravitational-wave modes during the inflationary epoch. Inflation moreover predicts that the square of the amplitude of these gravitational waves is proportional to the energy density $V(\phi)$ during inflation.

Such IGWs will produce temperature fluctuations in the CMB, primarily at large angular scales, with a spectrum that closely mimics the spectrum from density perturbations. It will thus be difficult to detect IGWs from the temperature pattern. However, as discussed further below, both density perturbations and gravitational waves produce polarisation in the CMB (Kamionkowski et al., 1997; Kamionkowski et al., 1997; Zaldarriaga & Seljak, 1997; Seljak & Zaldarriaga, 1997). When measured as a function of position on the sky, the Stokes parameters Q and U used to describe linear polarisation constitute components of a 2×2 symmetric trace-free tensor (Kamionkowski et al., 1997; Kamionkowski et al., 1997), or spin-2, field (Zaldarriaga & Seljak, 1997; Seljak & Zaldarriaga, 1997). Such a field can be decomposed into a curl-free (or longitudinal) component (an “E mode”) and a curl (or transverse) component (a “B mode”). Since primordial perturbations produce scalar perturbations to the spacetime metric, they can produce no curl. Gravitational waves, on the other hand, are tensor metric perturbations, and so they suffer no such restriction and can in fact produce a curl. The amplitude of this signal depends, of course, on the amplitude of the IGW background, and thus on the energy density during inflation. If inflation had something to do with grand unification, as many theorists might surmise, then the energy-density scale should be $V \sim (10^{15-16} \text{ GeV})^4$. If so, then the amplitude of the curl component is detectable by the kinds of instrumental sensitivities that should be available in forthcoming experiments (Kamionkowski & Kosowsky, 1998; Jaffe et al., 2000; Lewis et al., 2002; Kesden et al., 2002; Knox & Song, 2002; Seljak & Hirata, 2004). Detection

of such a curl component in the CMB polarisation would constitute a “smoking-gun” signal for inflation, and it has thus become a highly sought target for CMB experiments (BICEP¹, QUIET², QUAD³, POLARBEAR⁴) as well as an Einstein vision mission in NASA’s Beyond Einstein roadmap⁵.

Large-angle CMB temperature fluctuations from these gravitational waves (tensor metric perturbations) were first considered by Abbott & Wise (1984); Rubakov et al. (1982); Fabbri & Pollock (1983), and by Starobinsky (1985) while the polarisation was first considered by Polnarev (1985). Now, the most precise predictions for these power spectra come from numerical calculations (Seljak & Zaldarriaga, 1996). Like the power spectra for density perturbations (scalar metric perturbations), which exhibit wiggles due to acoustic waves in the primordial baryon-photon fluid, the temperature and polarisation power spectra from gravitational waves exhibit wiggles due to oscillations of tensor modes as they enter the horizon. The wiggles in the density-perturbation power spectra were predicted originally by Sunyaev & Zeldovich (1970) and Peebles & Yu (1970), and explained later elegantly with a semi-analytic approach in a paper by Hu & Sugiyama (1995).

The goal of this paper is to present an analytic account of the features in the tensor power spectra. Such an approach explains the origin of the features in the temperature and polarisation power spectra and illustrates the dependence of these features on the tensor power spectrum, cosmological parameters, and details of the recombination history. The intuition provided by such an approach complements the more precise results of numerical calculations. In particular, we explain here the location of the wiggles in the tensor temperature and polarisation power spectra, and why the bumps in the curl component of the polarisation are smoother than those in the curl-free component. We also show how the amplitude of the polarisation depends on the details of the recombination history. Our approach is analogous to that for scalar modes given by Hu & Sugiyama (1995). We discuss how measurement of the locations of these peaks can provide an independent probe of cosmological parameters.

The organisation of the paper is as follows. In Section 2.2, we write the exact equations for CMB fluctuations from tensor perturbations. The exact equations consist of the

¹<http://www.astro.caltech.edu/lgg/bicep/front.htm>

²<http://cfp.uchicago.edu/peterh/polarimetry/quiet3.html>

³http://www.stanford.edu/group/quest_telescope/

⁴<http://bolo.berkeley.edu/polarbear/>

⁵<http://universe.gsfc.nasa.gov/>

evolution of the gravitational waves, the visibility function, the source function, and projection factors. Next we develop a qualitative understanding of the physics contained in these relations in Section 2.3. The remainder of the paper then investigates individually each ingredient in the exact calculation. Section 2.4 discusses the evolution of the gravitational-wave perturbation. Section 2.5 discusses the effect of the recombination history on the power spectrum. Section 2.6 discusses the projection factors, and Section 2.7 the source function. Finally we comment on the dependence on cosmological parameters and detectability in Section 2.8. We include two Appendices that discuss the numerical techniques required to evolve the gravitational-wave amplitude in the presence of neutrino anisotropic stress (Appendix A), and the application of the WKB approach to gravitational waves evolving through the matter-radiation transition (Appendix A.2).

2.2 Exact equations

Here we present the exact equations required to evaluate the CMB power spectra from gravitational waves (Kamionkowski et al., 1997; Zaldarriaga & Seljak, 1997; Hu & White, 1997a). For simplicity, we will restrict ourselves to the case of a flat FRW universe. Our emphasis will be on small scale structure and so reionisation and its effects on large scales will not be discussed (see Zaldarriaga, 1997; Ng & Ng, 1996, for more details on this topic).

To provide the framework for temperature and polarisation anisotropies, we follow the formalism of Zaldarriaga & Seljak (1997). For two other useful introductions into the subject see Cabella & Kamionkowski (2003) and Lin & Wandelt (2006). The CMB radiation field is characterised by the Stokes parameters I , Q , and U . The intensity is described by I and polarisation along two axes at 45 degrees to one another by Q and U . The fourth Stokes parameter, V , which describes circular polarisation, is not generated by Thomson scattering, and while it can be generated after last scattering, the expected amplitudes are small (Cooray et al., 2003) and so can be neglected.

While convenient, the Stokes parameters Q and U describing polarisation suffer from being co-ordinate dependent. Under a right-handed rotation by an angle ψ in the plane perpendicular to the direction \hat{n} of propagation, Q and U transform according to

$$Q' = Q \cos 2\psi + U \sin 2\psi, \tag{2.1}$$

$$U' = -Q \sin 2\psi + U \cos 2\psi, \quad (2.2)$$

where $\hat{\mathbf{e}}'_1 = \cos \psi \hat{\mathbf{e}}_1 + \sin \psi \hat{\mathbf{e}}_2$ and $\hat{\mathbf{e}}'_2 = -\sin \psi \hat{\mathbf{e}}_1 + \cos \psi \hat{\mathbf{e}}_2$.

We wish to decompose the polarisation field described by Q and U in a basis set appropriate to the unit sphere and the above transformation properties. At first sight, it might seem appropriate to decompose Q and U using the usual spherical harmonics as a basis set. This would indeed be appropriate if Q and U were scalar functions of position on the unit sphere. However, a true scalar function would be invariant under rotation of the co-ordinate axes, which is clearly not the case for Q and U . Instead, it is fruitful to consider the combinations $Q \pm iU$, which transform under the above co-ordinate rotation as $(Q \pm iU) \rightarrow (Q \pm iU)e^{\mp 2i\psi}$. In the language of spin, these are quantities with spin-2, not the spin-0 of a scalar field. To properly represent these quantities, we follow Zaldarriaga & Seljak (1997) and turn to spin-weighted spherical harmonics, which generalise the symmetry properties of the spherical harmonics to quantities with spins other than spin-0. An alternative viewpoint is to consider Q and U as components of a second-rank tensor. This is the starting point for the formulation in Kamionkowski et al. (1997).

Making use of the spin-0 and spin-2 spin-weighted spherical harmonics $Y_{lm}(\hat{n})$ and ${}_{\pm 2}Y_{lm}(\hat{n})$ we can decompose the temperature and polarisation fields as

$$T(\hat{n}) = \sum_{lm} a_{T,lm} Y_{lm}(\hat{n}), \quad (2.3)$$

$$(Q \pm iU)(\hat{n}) = \sum_{lm} a_{\pm 2,lm} {}_{\pm 2}Y_{lm}(\hat{n}). \quad (2.4)$$

These expressions may be inverted to obtain the spherical-harmonic expansion coefficients,

$$a_{T,lm} = \int d\Omega Y_{lm}^*(\hat{n}) T(\hat{n}), \quad (2.5)$$

$$a_{\pm 2,lm} = \int d\Omega {}_{\pm 2}Y_{lm}^*(\hat{n}) (Q \pm iU)(\hat{n}). \quad (2.6)$$

Rather than work in terms of $a_{\pm 2,lm}$ it is advantageous to define two rotationally invariant quantities E and B by the relations

$$a_{E,lm} = -(a_{2,lm} + a_{-2,lm})/2, \quad (2.7)$$

$$a_{B,lm} = i(a_{2,lm} - a_{-2,lm})/2. \quad (2.8)$$

These two quantities are equivalent to the curl and grad modes defined in Kamionkowski et al. (1997). The E mode is invariant under the parity transformation, while the B mode transforms with odd parity.

From the above $a_{X,lm}$, we can form a series of correlation functions that characterise the statistics of the CMB perturbations. Of the six possible combinations, the TB and EB cross-correlations will vanish unless parity is somehow violated in the early Universe. The power spectra are defined as the rotationally-invariant quantities,

$$C_{XX'l} = \frac{1}{2l+1} \sum_m \langle a_{X,lm}^* a_{X',lm} \rangle. \quad (2.9)$$

Given a formalism to describe the observed CMB perturbations, it is then necessary to calculate a theoretical description of the perturbations. The starting point for this is to solve the Boltzmann equation for the radiation transfer of photons. To proceed, we expand the perturbations in Fourier modes of wavevector \mathbf{k} . A full derivation of the necessary equations is beyond the scope of this paper (for details see Zaldarriaga & Seljak, 1997; Polnarev, 1985; Crittenden, 1993), so we will summarise the important equations below.

For gravitational waves, the perturbed FRW metric takes the form

$$ds^2 = a(\tau)^2 \{-d\tau^2 + [\gamma_{ij} + 2h_{ij}(\mathbf{x}, \tau)] dx^i dx^j\}, \quad (2.10)$$

where γ_{ij} is the unperturbed flat space metric, and h_{ij} is traceless ($h^i_i = 0$) and transverse ($\partial_i h^{ij} = 0$), leaving two independent degrees of freedom corresponding to the two gravitational-wave polarisations. We will neglect the scalar and vector perturbations concentrating on the tensor h_{ij} perturbations. Note that we have written this line element in terms of the conformal time $\tau = \int^t dt'/a(t')$. We will use this temporal co-ordinate throughout and will denote with an overdot derivatives with respect to conformal time. The scale factor $a(\tau)$ is normalised to unity today.

With this co-ordinate choice, gravitational waves are represented by the transverse, traceless tensor metric perturbations h_{ij} . The evolution equations for the tensor modes

may be derived from the Einstein equations and are (Bertschinger, 1996; Bond, 1996)

$$\ddot{h}_{ij} + 2\frac{\dot{a}}{a}\dot{h}_{ij} + k^2 h_{ij} = 16\pi G a^2 \pi_{ij}. \quad (2.11)$$

Here, π_{ij} is the tensor part of the anisotropic stress and k is the comoving gravitational-wave wavenumber. For a mode travelling in the z direction, h_{ij} can be written in the form,

$$h_{ij} = \begin{pmatrix} h^+ & h^\times & 0 \\ h^\times & -h^+ & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

This shows explicitly how the two degrees of freedom describe two polarisation states h^+ and h^\times .

Tensor perturbations are assumed to arise, in similar fashion to scalar perturbations, from quantum fluctuations during inflation. Although our knowledge of this epoch is speculative, we may describe the statistical properties of the perturbations by a power spectrum $P_h(k)$ defined by,

$$\langle h^+(\mathbf{k})h^+(\mathbf{k}') \rangle = \langle h^\times(\mathbf{k})h^\times(\mathbf{k}') \rangle = \frac{P_h(k)}{2} \delta(\mathbf{k} - \mathbf{k}'). \quad (2.12)$$

If the process generating the perturbations is Gaussian, then this power spectrum encodes all information about the distribution. This primordial power spectrum is determined by inflation and can be written in terms of the Hubble parameter H evaluated at the time when CMB scales enter the horizon,

$$P_h(k) = \frac{32\pi G H^2}{(2\pi)^3 k^3} \Big|_{aH=k}. \quad (2.13)$$

In the case of slow-roll inflation, the value of H changes very little over the period when CMB scales enter the horizon, and it is appropriate to parameterise the power spectrum in terms of an amplitude A_T , fixed by the energy scale of inflation, and a tensor spectral index n_T ,

$$P_h(k) = A_T k^{n_T-3}. \quad (2.14)$$

Slow-roll inflationary theories generically predict the tensor spectral index $n_T \approx 0$, a nearly

scale-invariant spectrum.

The temperature and polarisation anisotropies induced by an equal mixture of tensor modes of + and \times polarisation with amplitude h may be described in terms of the variables $\Delta_X(\tau_0, \hat{\mathbf{n}}, \mathbf{k})$, where $X=(T,E,B)$. The gravitational wave causes distortions with an angular pattern that depends on both $\mu = \hat{\mathbf{n}} \cdot \hat{\mathbf{k}}$, the angle between the direction $\hat{\mathbf{n}}$ of propagation of the photon, and the wavevector $\hat{\mathbf{k}}$ of the tensor mode, and on the azimuthal angle ϕ . Polnarev (1985) was the first to recognise the advantage of separating out this angular dependence by introducing new variables $\tilde{\Delta}_T$ and $\tilde{\Delta}_P$ defined by (Polnarev, 1985; Kosowsky, 1996)

$$\Delta_T^+(\tau, k, \mu, \phi) = (1 - \mu^2) \cos(2\phi) \tilde{\Delta}_T^+(\tau, k, \mu), \quad (2.15)$$

$$\Delta_Q^+(\tau, k, \mu, \phi) = (1 + \mu^2) \cos(2\phi) \tilde{\Delta}_P^+(\tau, k, \mu), \quad (2.16)$$

$$\Delta_U^+(\tau, k, \mu, \phi) = -2\mu \sin(2\phi) \tilde{\Delta}_P^+(\tau, k, \mu), \quad (2.17)$$

and similar equations for the \times polarisation with $\cos(2\phi) \rightarrow \sin(2\phi)$ and $\sin(2\phi) \rightarrow \cos(2\phi)$. The ϕ dependence reflects the intrinsic angular dependence of the gravitational waves, while the μ dependence is chosen to simplify the evolution equations. For our purposes, we may use these Polnarev variables for calculating the sources; the sources are then simply related to the original variables (Zaldarriaga & Seljak, 1997). The derivations of these evolution equations and relations is beyond the scope of this paper, so we simply quote the results.

The evolution of a single Fourier mode k satisfies the Boltzmann equations,

$$\dot{\tilde{\Delta}}_T + ik\mu\tilde{\Delta}_T = -\dot{h} - \dot{\kappa}[\tilde{\Delta}_T - \Psi], \quad (2.18)$$

$$\dot{\tilde{\Delta}}_P + ik\mu\tilde{\Delta}_P = -\dot{\kappa}[\tilde{\Delta}_P + \Psi], \quad (2.19)$$

$$\Psi \equiv \left[\frac{1}{10}\tilde{\Delta}_{T0} + \frac{1}{7}\tilde{\Delta}_{T2} + \frac{3}{70}\tilde{\Delta}_{T4} - \frac{3}{5}\tilde{\Delta}_{P0} + \frac{6}{7}\tilde{\Delta}_{P2} - \frac{3}{70}\tilde{\Delta}_{P4} \right]. \quad (2.20)$$

Here, we have defined the differential cross section for Thomson scattering as $\dot{\kappa} = an_e x_e \sigma_T$, where n_e is the electron number density, x_e is the ionisation fraction, and σ_T is the Thomson cross section. The total optical depth between a conformal time τ and τ_0 is given by integrating $\dot{\kappa}$ to obtain $\kappa(\tau, \tau_0) = \int_{\tau}^{\tau_0} \dot{\kappa}(\tau) d\tau$. The multipole moments of temperature and of polarisation are defined by $\Delta(k, \mu) = \sum_l (2l + 1)(-i)^l \Delta_l(k) P_l(\mu)$, where $P_l(\mu)$ is the Legendre polynomial of order l . This decomposition converts Eqs. (2.18) and (2.19) into

an infinite hierarchy of equations connecting higher moments to lower moments. These equations have solutions (Zaldarriaga & Seljak, 1997)

$$\Delta_{Tl} = \sqrt{\frac{(l+2)!}{(l-2)!}} \int_0^{\tau_0} d\tau S_T(k, \tau) \frac{j_l(x)}{x^2}, \quad (2.21)$$

$$\Delta_{El} = \int_0^{\tau_0} d\tau S_P(k, \tau) \left[-j_l(x) + j_l''(x) + \frac{2j_l(x)}{x^2} + \frac{4j_l'(x)}{x} \right], \quad (2.22)$$

$$\Delta_{Bl} = \int_0^{\tau_0} d\tau S_P(k, \tau) \left[2j_l'(x) + \frac{4j_l(x)}{x} \right], \quad (2.23)$$

where $j_l(x)$ is the spherical Bessel function. In these expressions, $x = k(\tau_0 - \tau)$. Defining the visibility function $g(\tau) = \kappa e^{-\kappa}$, the sources are given by

$$S_T(k, \tau) = -\dot{h}e^{-\kappa} + g\Psi, \quad (2.24)$$

$$S_P(k, \tau) = -g\Psi, \quad (2.25)$$

and the power spectra by

$$C_{XX'l} = (4\pi)^2 \int k^2 dk P_h(k) \Delta_{Xl}(k) \Delta_{X'l}(k). \quad (2.26)$$

It is straightforward to show that for statistically equal distributions of left and right circularly polarised gravitational waves, the TB and EB cross-correlations vanish. If there is a preference for either polarisation, then a non-zero TB and EB correlation will be observed (Lue et al., 1999).

Equations for the evolution of the ionisation fraction, and hence the optical depth, exist, but will not be dealt with here. Details of the relevant equations and useful analytic approximations for the optical depth and ionisation fraction may be found in Jones & Wyse (1985).

Where necessary, we assume a fiducial Λ CDM cosmology with $\Omega_b = 0.05$, $\Omega_{DM} = 0.25$, $\Omega_\Lambda = 0.7$, and with Hubble parameter parameterised by $h = 0.72$. When tensor power spectra are plotted, we have used the normalisation $A_T = 2.4 \times 10^{-11}$, corresponding to an inflationary energy scale $V^{1/4} = 2 \times 10^{16}$ GeV, and tensor spectral index $n_T = 0$. We

then plot the dimensionless quantity $l(l+1)C_l/2\pi$, which gives the fractional variance per logarithmic interval in l .

The above set of equations forms the basis for our problem. Having written expressions for the power spectra, we must now exploit a mixture of physical and mathematical approximations to bring out their implications.

2.3 A tale of tensor modes

Let us now try to obtain an intuitive understanding of how the features in the power spectra arise. This will help motivate the approximations that follow in later sections. Useful discussions of how polarisation is generated are given in Cabella & Kamionkowski (2003); Dodelson (2003), and in Hu & White (1997b).

First let us discuss the temperature power spectrum. The temperature multipole moments due to an individual gravitational wave of wavenumber k observed at a conformal time τ_0 are

$$\Delta_{Tl} = \sqrt{\frac{(l+2)!}{(l-2)!}} \int_0^{\tau_0} d\tau \left(-\dot{h}e^{-\kappa} + g\Psi \right) \frac{j_l(x)}{x^2}, \quad (2.27)$$

with $x = k(\tau_0 - \tau)$. The second of the sourcing terms is localised to the surface of last scattering (SLS) by the visibility function; as a consequence of the restricted range this term is small and may be neglected at all angular scales. Between $l = 200$ and $l = 800$, the contribution from the $g\Psi$ term falls off more slowly than the integrated Sachs-Wolfe (ISW) term allowing it to become marginally relevant. At lower and higher l , the power generated by the second term dies off rapidly and is totally negligible (Fig. 2.1). The first term, which dominates this integral, involves an integral from the SLS to the present day. Its form tells us that the temperature power spectrum is sensitive to the evolution of gravitational waves from the SLS to today and not to the recombination history. This term is a form of integrated Sachs-Wolfe (ISW) effect that describes the anisotropy generated by changing gravitational potentials. We can understand this effect by recalling that a gravitational wave alternately stretches and compresses space as it oscillates. A photon travelling past the gravitational wave loses energy when its wavelength is stretched, but gains energy when its wavelength is reduced. If the gravitational-wave amplitude evolves over the course of the oscillation, the photon will undergo a net change in energy. Tensor

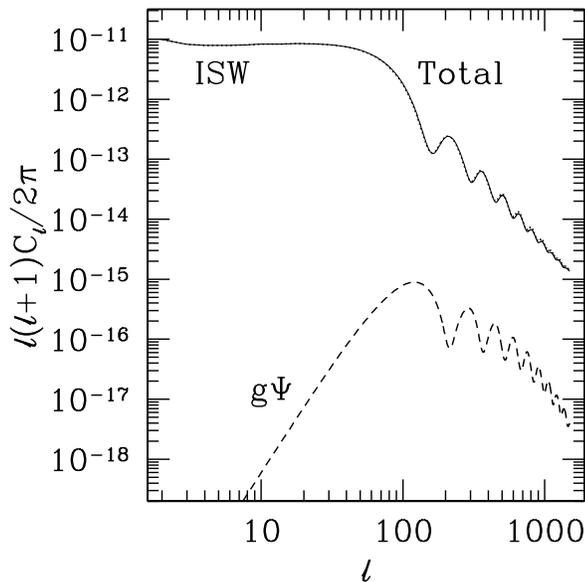


Figure 2.1: Comparison of power generated by the two source terms for temperature anisotropy. Plotted are the total power (solid line), ISW term only (dotted line), and $g\Psi$ term only (dashed line). The $g\Psi$ term is essentially negligible at all l . For this and all other power-spectra plots, we have used the normalisation $A_T = 2.4 \times 10^{-11}$, corresponding to an inflationary energy scale $V^{1/4} = 2 \times 10^{16} \text{GeV}$, and tensor spectral index $n_T = 0$.

modes decrease steadily in amplitude and oscillate after horizon entry. As such, a photon travelling along the crest of a phase front will slowly gain energy as the overall amplitude of the gravitational wave decreases. Photons travelling at an angle to the mode experience further red and blue shifting as they propagate through different phase regions. Their energy oscillates as a consequence. Between the SLS and today, the period-averaged amplitude of the tensor mode decreases, and so the mean energy of the photon increases.

If we consider a late time, so that the amplitude of the tensor modes is essentially zero, then we see that the final energy of the photon is determined by whether it started its journey from the SLS at a trough or crest in the tensor mode. Photons starting at a crest will have gained more energy and appear hotter than average and vice versa for those starting at a trough. This simplistic picture is modified by the effect of power free-streaming from one angular scale to another as the Universe expands, which tends to smooth the resultant power spectra.

The situation is very different for the polarisation anisotropies. These are generated by

expressions of the form

$$\Delta_{Xl} = \int_0^{\tau_0} d\tau (-g\Psi) P_{Xl}[k(\tau_0 - \tau)]. \quad (2.28)$$

Here, the source is very firmly localised to the SLS and so is sensitive to the thermal history and gravitational-wave evolution at that time. This is sensible. Treating the early radiation bath as unpolarised (as we expect from suppression of anisotropy during the tightly-coupled regime), then polarisation is generated by Thomson scattering of an anisotropic intensity distribution. Where does this anisotropy come from? In the rest frame of the scattering electron, photons arrive from all directions from a mean distance determined by the mean free path of photons at recombination. In propagating, these photons experience the ISW effect, discussed in the case of the temperature spectrum, and so arrive at the scatterer with altered temperatures. The resulting anisotropic temperature distribution is scattered, generating polarisation which free-streams to the present epoch.

In this way, we can understand the power spectrum. For modes with wavelengths much larger than the horizon size at last scattering, incident photons experience very little ISW before the last scattering event and little polarisation is generated. Optimal ISW and thus maximal polarisation is generated by modes that enter the horizon at the time of penultimate scattering. The amplitude of the gravitational wave decays most rapidly immediately on horizon entry (see Fig. 2.2) before settling into oscillation with a slowly decreasing amplitude. Modes that enter the horizon before penultimate scattering lead to photons whose ISW samples this slowly decaying regime. Hence, they generate significant polarisation, but less than for the optimal case. Note that the time between penultimate and last scattering will be about the width of the surface of last scattering.

Translating this into the form of the polarisation power spectrum, we expect a slow increase in power at large scales peaking at the scale of the horizon at penultimate scattering. Immediately after this, we expect a large drop in polarisation corresponding to the transition between modes that enter the horizon between penultimate and last scattering and those that do not. Next, we expect a steady decline in power as modes have entered the horizon before penultimate scattering and so redshifted away before the ISW effect is generated. This region will show a transition in slope between modes that entered the horizon in the matter- and radiation-dominated epochs. On scales smaller than the mean free path at

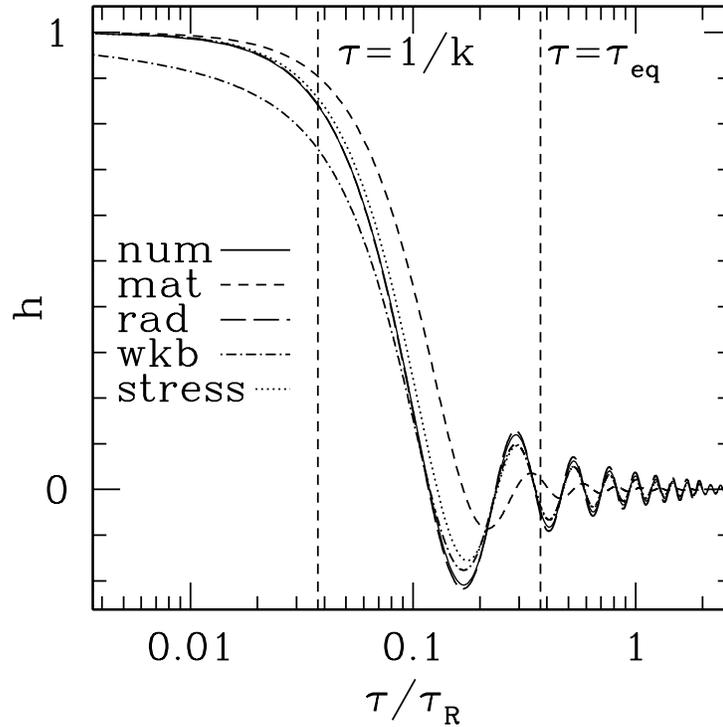


Figure 2.2: Evolution of the gravitational-wave amplitude h for wavenumber k satisfying $k\tau_{\text{eq}} = 10$. The plots are normalised so that $h(\tau = 0) = 1$. Shown results are numerical solution of Eq. (2.29) without anisotropic stress (solid curve), numerical solution of Eq. (2.29) with anisotropic stress (dotted curve), radiation dominated (Eq. (2.31), long dashed curve), matter dominated (Eq. (2.32), short dashed curve), and WKB (Eq. (2.36), dot-dashed curve). The two vertical lines denote $\tau = 1/k$ and τ_{eq} .

recombination, the power will drop sharply as phase cancellation between differing crests and troughs becomes important.

In this discussion, it is important to realise that only three scales have entered the problem. These are the comoving horizon at recombination, the horizon at matter-radiation equality, and the width of the last-scattering surface. Fig. 2.3 shows how the features in the power spectrum correspond to these scales.

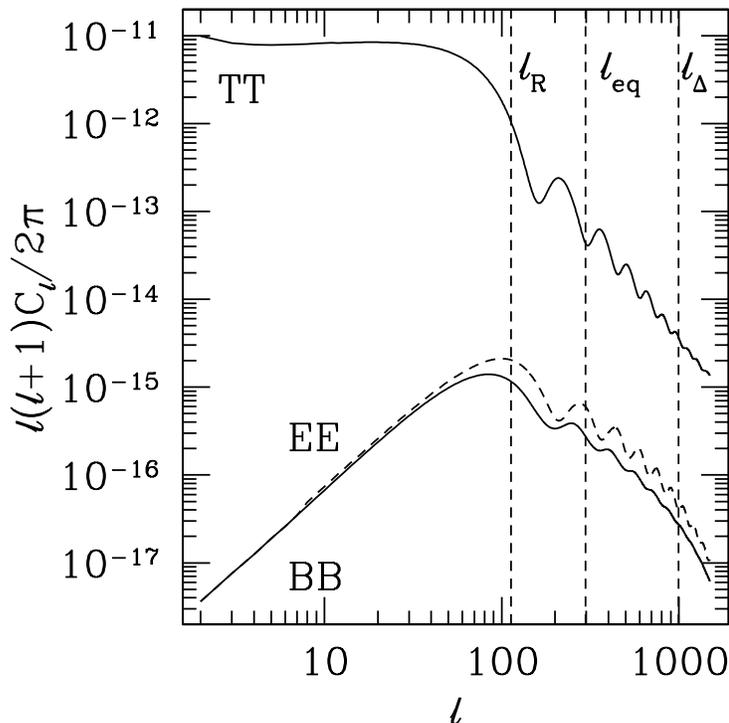


Figure 2.3: Tensor power spectra. Curves from top to bottom are C_l^{TT} , C_l^{EE} , and C_l^{BB} . Vertical lines indicate important angular scales, from left to right: horizon at recombination, τ_R , horizon at matter-radiation equality, τ_{eq} , and the width of the last-scattering surface, $\Delta\tau_R$.

2.4 Gravitational-wave evolution

Expansion of the Universe leads to damping of the tensor modes as described by the term proportional to \dot{h}_{ij} in Eq. (2.11). This is the usual redshifting of radiation. In addition, the tensor modes may be sourced by anisotropic stress, π_{ij} . It has been shown (Bond, 1996; Weinberg, 2004) that anisotropic stress generated by free-streaming neutrinos acts to

provide viscosity, further damping the tensor modes. This effect is important only while the energy density in neutrinos is a significant fraction of the total energy; i.e., during the radiation-dominated epoch.

The tensor modes may be decomposed into two independent polarisation states, h_{\times} and h_{+} . With this decomposition and a source term appropriate for neutrino anisotropic stress, we have

$$\ddot{h}_i + 2\frac{\dot{a}}{a}\dot{h}_i + k^2 h_i = -24f_{\nu}(\tau) \left(\frac{\dot{a}(\tau)}{a(\tau)}\right)^2 \int_0^{\tau} K[k(\tau - \tau')] \dot{h}_i(\tau') d\tau', \quad (2.29)$$

where $i = +, \times$, and $f_{\nu} \equiv \bar{\rho}_{\nu}/\bar{\rho}$ with $\bar{\rho}$ the unperturbed density, and $K(s)$ is given by

$$K(s) \equiv -\frac{\sin s}{s^3} - \frac{3 \cos s}{s^4} + \frac{3 \sin s}{s^5}. \quad (2.30)$$

To a first approximation, we may neglect the effect of anisotropic stress, though it should be included in detailed calculations. Without the source term, analytic solutions for Eq. (2.29) in pure radiation and matter cosmologies may be expressed in terms of the spherical Bessel function $j_l(x)$,

$$h_{\text{rad}}(\tau) = h(0)j_0(k\tau) = h(0)\frac{\sin k\tau}{k\tau}, \quad (2.31)$$

$$h_{\text{mat}}(\tau) = 3h(0)\frac{j_1(k\tau)}{k\tau}. \quad (2.32)$$

In a mixed radiation and matter dominated universe, the solution follows h_{rad} initially before asymptotically becoming similar to h_{mat} . The initial radiation dominated phase introduces a phase shift into h_{mat} as now the boundary conditions do not preclude the spherical Neumann solution to the unsourced Eq. (2.29). When calculating the power spectra it is important to get this phase, which determines the peak positions, correct. This point was understood but not implemented in Turner et al. (1993) and included implicitly by others (Wang, 1996).

The behaviour of these solutions is shown in Fig. 2.2 and splits into three main regimes. When $k\tau \ll 1$, h evolves slowly and is approximately constant. Once $k\tau \approx 1$, the amplitude decays away rapidly before entering an oscillatory phase with slowly decreasing amplitude, when $k\tau \gg 1$. Physically, this corresponds to a mode that is frozen beyond the horizon until its wavelength is of order the horizon size at which point it enters the horizon and redshifts rapidly with the expansion of the Universe.

Recombination occurs shortly after the Universe becomes matter dominated. For modes that enter the horizon during the matter-dominated regime and so have evolved little in the radiation-dominated epoch, we expect h_{mat} to be a good description. For modes that entered during radiation domination, we expect that the transition from radiation to matter domination will affect the evolution significantly.

The matter-radiation transition can be accounted for in a variety of ways. Most simple is to assume that the transition is instantaneous and to match the amplitude and derivative of h on the boundary. This will be a good approximation for waves with wavelength much longer than the time taken for the transition to take place.

$$h_{\text{instant}} = \begin{cases} j_0(k\tau), & \tau < \tau_{\text{eq}}, \\ (\tau_{\text{eq}}/\tau)[Aj_1(k\tau) + By_1(k\tau)], & \tau > \tau_{\text{eq}}, \end{cases} \quad (2.33)$$

with

$$A = \frac{\frac{3}{2}k\tau_{\text{eq}} - \frac{1}{2}k\tau_{\text{eq}} \cos(2k\tau_{\text{eq}}) + \sin(2k\tau_{\text{eq}})}{k^2\tau_{\text{eq}}^2}, \quad (2.34)$$

$$B = \frac{2 - 2k^2\tau_{\text{eq}}^2 - 2\cos(2k\tau_{\text{eq}}) - k\tau_{\text{eq}} \sin(2k\tau_{\text{eq}})}{2k^2\tau_{\text{eq}}^2}. \quad (2.35)$$

Alternatively, we may consider the situation where the wavelength of the gravitational wave is much shorter than the transition time. In this case, the gravitational wave sees the background expansion vary slowly and a WKB approach is appropriate. Ng & Speliotopoulos (1995) first presented this approach, although they were primarily interested in late time asymptotic limits and so neglected the behaviour near the classical turning point. Here we generalise their result making use of the uniform Langer solution for the WKB problem (Bender & Orszag, 1978). The result is

$$h(\tau) = \frac{\Gamma(k\tau)^{-1/4}}{\tau^{1/2}(\tau+2)} \left(\frac{3}{2}S_0(\tau)\right)^{1/6} \times \left\{ 2\sqrt{\pi}C_2\text{Ai} \left[\left(\frac{3}{2}S_0(\tau)\right)^{2/3} \right] + \sqrt{\pi}C_1\text{Bi} \left[\left(\frac{3}{2}S_0(\tau)\right)^{2/3} \right] \right\}, \quad (2.36)$$

with

$$\Gamma(s) = \frac{1}{4} + \frac{2s}{s+2k} - s^2, \quad (2.37)$$

and

$$S_0(\tau) = \int_{k\tau}^{k\tau_T} \sqrt{\Gamma(s)} \frac{ds}{s}. \quad (2.38)$$

Here, τ_T is the solution to $\Gamma(k\tau) = 0$, Ai and Bi are Airy functions, and C_1 and C_2 are constant coefficients set by the boundary conditions $h(0) = 1$ and $\dot{h}(0) = 0$. For technical reasons, these boundary conditions must be extrapolated to small τ via asymptotic approximation to Eq. (2.29) and then applied. Care must be taken in evaluating the above expressions when $\tau > \tau_T$. These details are discussed further in Appendix A.2. This WKB expression reproduces the phase of h in both radiation and matter dominated regimes, but underestimates the amplitude. The close agreement between the WKB and anisotropic-stress curves in Fig. 2.2 is a numerical coincidence.

Other approaches exist to handle this transition from radiation to matter in a more pragmatic fashion (Turner et al., 1993; Wang, 1996).

We can get the scaling of h from a simple argument. Before horizon entry, the amplitude h of a gravitational wave is constant. After horizon entry, the gravitational wave redshifts with the expansion as radiation and scales as $h \sim 1/a$. Hence, the amplitudes of a gravitational wave today and at horizon entry are related by $h_{\text{today}}/h_{\text{entry}} = a_{\text{entry}}/a_{\text{today}}$. Taking h_{entry} to be independent of k , we have $h_{\text{today}} \propto a_{\text{entry}}$. Horizon entry occurs when $a_{\text{entry}} H_{\text{entry}} = k$, and so from the scaling of H in the matter- and radiation-dominated epochs we obtain $a_{\text{entry}} \propto k^{-1}$ when radiation dominated and $a_{\text{entry}} \propto k^{-2}$ when matter dominated. Thus, we obtain the scalings,

$$h \propto \begin{cases} 1, & k < 1/\tau_0, \\ k^{-2}, & 1/\tau_{\text{eq}} > k > 1/\tau_0, \\ k^{-1}, & k > 1/\tau_{\text{eq}}. \end{cases} \quad (2.39)$$

This result agrees with both the instantaneous-transition and WKB solutions when $\tau \gg \tau_{\text{eq}}$. These scaling relations form the basis for scaling of the power spectrum. We expect $l(l+1)C_l^{TT}$ to scale as (Starobinsky, 1985; Turner et al., 1993)

$$l(l+1)C_l^{TT} \propto \begin{cases} 1, & l < l_R, \\ l^{-4}, & l_{\text{eq}} > l > l_R, \\ l^{-2}, & k > l_{\text{eq}}. \end{cases} \quad (2.40)$$

It has been claimed (Starobinsky, 1985; Atrio-Barandela & Silk, 1994) that there should be an extra region, $l > l_\Delta$, in which the width of the last-scattering surface becomes important and due to phase-damping the scaling goes as l^{-6} . However the dominant source of temperature anisotropy is the ISW effect, which is insensitive to the recombination history, and so we do not expect to see this behaviour in the temperature power spectrum. One way to see this is to examine the kernel in Eq. (2.21). On small scales, which enter the horizon before τ_R , the finite rise time of $e^{-\kappa}$ alters the weight in the integral by a nearly constant factor. For all other modes, it is sufficient to simply truncate the range of the integral to between τ_R and τ_0 , effectively imposing instantaneous recombination. On the other hand, the polarisation anisotropy is generated near the SLS and will show phase cancellation dependent on the width of the SLS. We will return to this point later.

In the absence of reionisation, at low l the power spectra for the polarisation grow as l^2 (Hu & White, 1997a). From this and the above scaling arguments, we would expect the power spectrum to scale as

$$l(l+1)C_l^{XX} \propto \begin{cases} l^2, & l < l_R, \\ l^{-2}, & l_{\text{eq}} > l > l_R, \\ 1, & l_{\text{eq}} < l < l_\Delta, \\ l^{-4}, & l > l_\Delta, \end{cases} \quad (2.41)$$

with X=(E,B). The effect of phase-damping extends to much lower l than would be indicated by these simple dimensional arguments. Consequently, the region of constant power $l_{\text{eq}} < l < l_\Delta$ is never visible in calculated spectra, but is lost in the transition to the phase-damping regime.

The above expressions for the gravitational-wave amplitude are used to generate the power spectra displayed in Fig. 2.4. The plots are normalised by taking $n_T = 0$ and setting $A_T = 1$. All of the plots show the same scaling relation at low l . This regime is dominated by modes that have not entered the horizon at recombination and so are approximately constant. At low l h_{mat} underestimates the power, while h_{rad} overestimates the power. This is a consequence of the contribution of the modes that have entered the horizon that are evolving in a mixed radiation-matter universe and so have amplitudes intermediate to the predictions of these two approximations.

Moving above the peak at l_R , we clearly see the different scaling relations between the matter and radiation approximations. Recall that at these high l s, we expect the main contribution to come from modes that entered the horizon in the radiation-dominated epoch, and so h_{rad} should be a good approximation. The WKB result shows a transition between following the matter-dominated curve to behaving more like the radiation-dominated form, though with reduced amplitude. This reduction in amplitude is an unfortunate characteristic of the WKB solution and is not significant to understanding the physics. The WKB solution serves as a nice bridge between matter- and radiation-dominated epochs. The instantaneous solution fails to be useful on scales with wavelength short compared to the transition time-scale.

These curves display the scaling expected from Eq. (2.39), but we see that in the numerical case, excluding anisotropic stress, we never observe the full l^{-4} scaling for a matter-dominated regime. The combination of recombination occurring soon after matter-radiation equality and the Universe becoming matter dominated only slowly means that the power spectra damp more slowly, closer to l^{-3} , for scales $1/\tau_R < k < 1/\tau_{\text{eq}}$. The presence of matter also causes peak positions to shift to smaller scales over the fully radiation-dominated case indicative of the phase shift that the transition introduces in h .

For the purposes of reproducing the exact tensor-mode power spectra, we must worry about preserving both the amplitude and phase of the gravitational waves. The importance of the amplitude is clear in estimating the power correctly. The phase determines the positions of the maxima and minima in the high- l region of the spectra. Maxima correspond to gravitational waves whose amplitude was at a maximum or minimum at the SLS; minima correspond to gravitational waves whose amplitude was close to zero at the SLS. Altering the phase of the gravitational waves shifts the k values for which these maxima occur at the SLS and so shift the features in the CMB. If we wish to understand these features in detail, then we must understand how the phase of the gravitational waves varies with k and how this is mapped onto the power spectrum. This mapping is the subject of Section 2.6.

2.5 Recombination history

While the Universe is young and hot, baryons are ionised and tightly coupled to photons via Thomson scattering. Once the temperature falls below a few eV, it becomes favourable

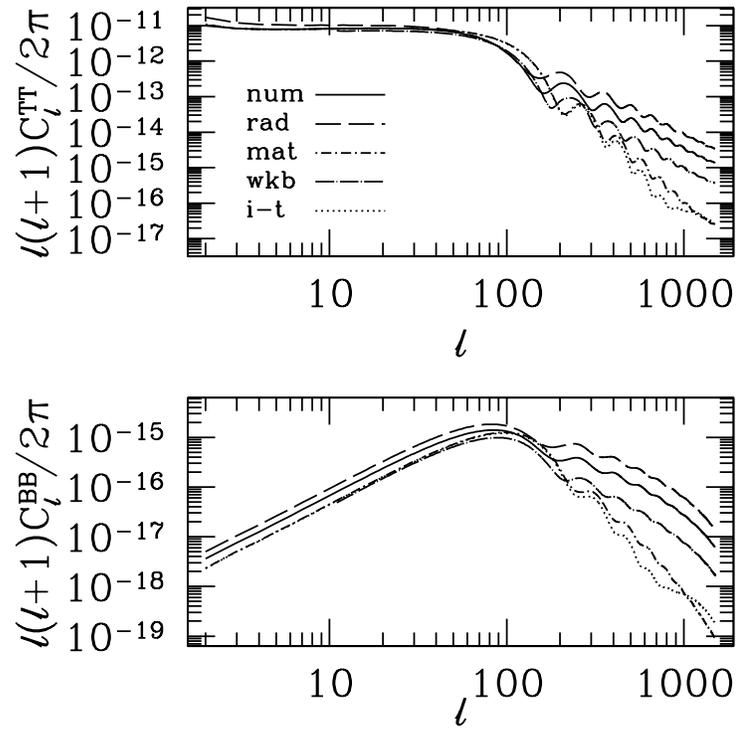


Figure 2.4: T and B power spectra calculated using approximate forms for the gravitational-wave amplitude h . Plotted are the results using h from the full numerical calculation without anisotropic stress (solid curve) and from the radiation-dominated (long dashed curve), matter-dominated (dot-short dashed curve), instantaneous-transition (dotted curve), and the WKB (dot-long dashed curve) approximations.

for electrons and ions to recombine to form neutral molecules. As the number of charged particles falls, the mean free path of any given photon increases. Eventually, the mean free path becomes comparable to the horizon size and the photon and baryon fluids are essentially decoupled. It is at this point in the Universe's evolution that the CMB photons last scatter.

The visibility function describes the probability that a given CMB photon last scattered from a particular time. In terms of the optical depth κ , this visibility function is given by

$$g(\tau) = \kappa e^{-\kappa}. \quad (2.42)$$

Numerical calculations show that $g(\tau)$ is sharply peaked during recombination. This property suggests we approximate the visibility function by a narrow Gaussian for analytic simplicity. For example,

$$g(\tau) = g(\tau_R) e^{-\frac{(\tau - \tau_R)^2}{2\Delta\tau_R^2}}, \quad (2.43)$$

determines the visibility function in terms of the conformal time τ_R of recombination, its width $\Delta\tau_R$, and the amplitude $g(\tau_R)$ at recombination.

Approximating the visibility by a Gaussian leads to a simple form for the optical depth in the region close to τ_R . If we write κ in the general form $\kappa = \exp[-f(\tau)]$, then consistency with Eqs. (2.42) and (2.43) requires that $\kappa \approx \exp[-(\tau - \tau_R)/\Delta\tau_R]$ and $g(\tau_R) \approx 1/(e\Delta\tau_R)$ in the region close to τ_R . This latter result is essentially a statement about the normalisation of the Gaussian and preserves the total weight of the visibility function for different widths. Away from recombination, the evolution of the optical depth is a complicated function of the thermal history and not easily approximated.

These approximations for $g(\tau)$ and κ are plotted in Fig. 2.5 for the fiducial cosmology with $\Omega_b = 0.05$, $\Omega_{\text{DM}} = 0.25$, and $\Omega_\Lambda = 0.7$. For this cosmology, we have $\tau_0 = 13515 \text{ Mpc}$, $\tau_R/\tau_0 = 0.0203$, $\tau_{\text{eq}}/\tau_0 = 0.0076$, and $\Delta\tau_R/\tau_0 = 0.0012$. While the Gaussian form does a reasonable job of approximating the shape of the peak, the visibility function is clearly skewed and possesses a significant tail. The combination of these features means the Gaussian approximation will underestimate the power and shift features to slightly smaller angles than in the true power spectrum.

Fig. 2.6 shows a series of power spectra calculated using the Gaussian approximation.

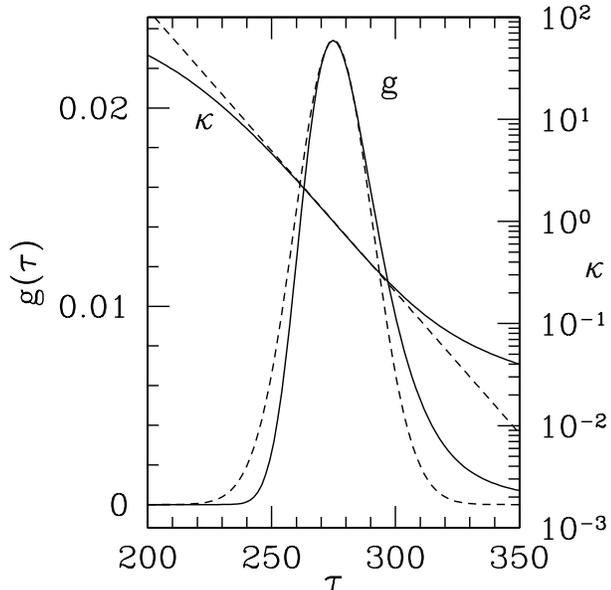


Figure 2.5: Recombination history. Plotted are the visibility function $g(\tau)$ and the optical depth κ calculated numerically using CMBFAST (solid curves) and the approximations described in Eq. (2.43) and the text (dashed curves) using $\Delta\tau_R = 15.7$.

In each, the correct thermal history is used to calculate the evolution of the source function Ψ with the Gaussian approximation applied when calculating the Δ_X from Eqs. (2.24) and (2.25). Although not strictly self-consistent, this isolates the modification of the source due to a changed thermal history from the effect of the visibility function on generating anisotropies. Source evolution will be considered in Section 2.7.

The temperature power spectrum shows no variation with $\Delta\tau_R$ at $l < 200$. Power on these scales is generated via the integrated Sachs-Wolfe effect by modes that only evolve significantly between τ_R and τ_0 and so are insensitive to the thermal history. At smaller scales, the modes of interest are evolving over recombination and so contain information about the thermal history. Modifying the width of the visibility function affects the power spectrum via the $e^{-\kappa}$ term in Eq. (2.27) which acts to cut the integral off below τ_R . Widening the SLS makes this cutoff slower which, owing to the concave nature of $e^{-\kappa}$, leads to less weight in the integral. This leads to the differences observed in the top panel of Fig. 2.6. This is not phase-damping, and does not alter the scaling of the power spectrum significantly. In addition to this overall shift in power, larger $\Delta\tau_R$ acts to wash out the bumps and wiggles. For modes that oscillate rapidly over this rise time, the ISW samples

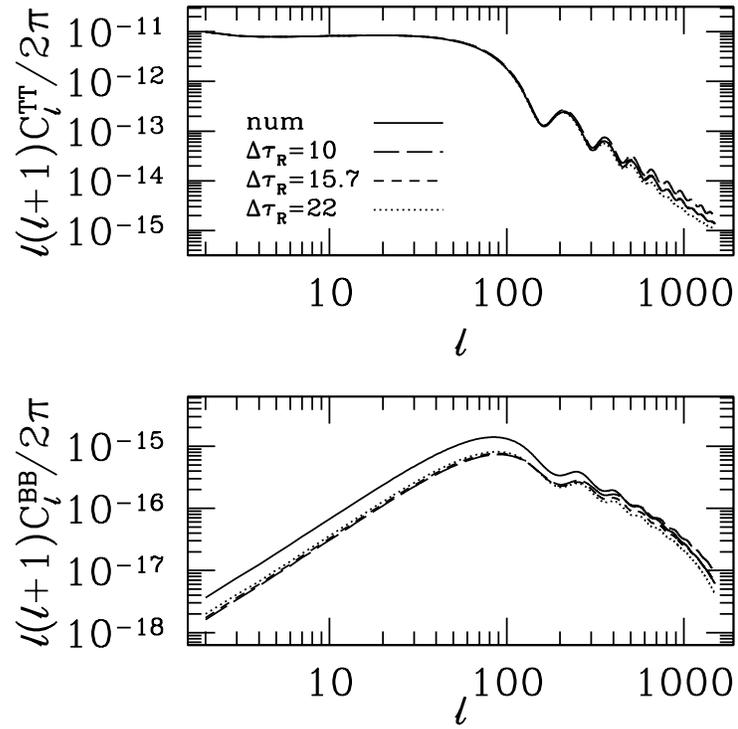


Figure 2.6: Evaluation of using the Gaussian approximation for the visibility function on the power spectra. Three power spectra calculated using Eq. (2.43) for the visibility function are shown for values of $\Delta\tau_R = 10$ (long dashed curve), 15.7 (short dashed curve), and 22 (dotted curve).

an averaged starting value of h and so is less sensitive to the presence of peaks and troughs. Power is still generated from the net decrease in the maximum amplitude of h from τ_R to τ_0 , so this effect does not cause a large drop in power as suggested in Starobinsky (1985) and in Atrio-Barandela & Silk (1994).

Fig. 2.6 shows that the Gaussian approximation leads to a lower polarisation power spectrum. This difference is a consequence of the long tail to the visibility function, which is not reproduced in the Gaussian approximation. In these plots, varying $\Delta\tau_R$ does not affect the overall amplitude of the power spectrum. This is an artefact of using the same Ψ for each plot. In reality, the amplitude of the polarisation power spectra depends sensitively on $\Delta\tau_R$, as will be shown in Section 2.7. These plots show that varying $\Delta\tau_R$ in the Gaussian approximation does not affect the shape at low l , but a wider width leads to a sharper fall off in power at high l . This is a feature of phase-damping, which will be discussed in Section 2.7.

None of the three values used precisely reproduces the decline of the true power spectrum, which is seen to fall off more rapidly than the approximations. This seems to be a consequence of the tail to the visibility function. In keeping with expectation, the peaks in the high- l region are found at slightly higher l in the approximations than the numerical result.

2.6 Projection factors

The power spectra that we observe today are projections of the temperature and polarisation anisotropies at the last-scattering surface. By inspection of Eqs. (2.21), (2.22), and (2.23) we can define three projection terms,

$$P_{Tl}(x) = \frac{j_l(x)}{x^2}, \quad (2.44)$$

$$P_{El}(x) = -j_l(x) + j_l''(x) + \frac{2j_l(x)}{x^2} + \frac{4j_l'(x)}{x}, \quad (2.45)$$

$$P_{Bl}(x) = 2j_l'(x) + \frac{4j_l(x)}{x}. \quad (2.46)$$

Typically the argument of these terms is $k(\tau_0 - \tau)$, the look-back time scaled by the wavenumber, reflecting that these are projections from the point of origin onto today's

sky.

The different forms of the projection factors, plotted in Fig. 2.7, help explain many of the features seen in the power spectra (Fig. 2.8).

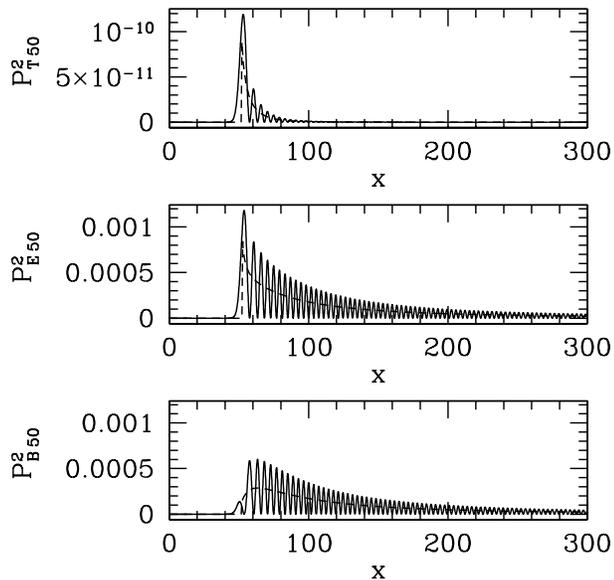


Figure 2.7: Comparison of exact and approximate projection terms evaluated for $l = 50$. Plotted are the exact forms from Eqs. (2.44), (2.45), and (2.46) (solid curves) and the approximate forms from Eqs. (2.48), (2.49), and (2.50) (dashed curves).

First consider the E projection factor as a simple example. A sharp peak occurs at $x \approx l$. This tells us that the value of C_l^{TT} at $l = 50$ is determined by the behaviour of the source function at $x \approx l$. The polarisation source function is strongly peaked around $\tau = \tau_R$, which implies the behaviour at last scattering of the mode with wavenumber $k \approx l/(\tau_0 - \tau_R)$ dominates the contribution to C_l . If the projection factor was a Dirac delta function, this would be the whole story. However, the projection factor has a significant tail for $x > l$ signifying that modes with larger wavenumber also contribute power to this angular scale.

From this, we can see that the sharper the spike at $x \approx l$, the sharper the features seen in the power spectrum. A wider peak mixes in modes of different phases blurring the spectra. Noting that the B projection factor lacks a sharp peak, we expect the B power spectra to contain blurred features relative to the E spectra. In addition, as its maximum is at higher x , we expect features in the source to be shifted to smaller l than in the E spectrum. This sort of argument has some validity with the T spectrum, but is there complicated by the

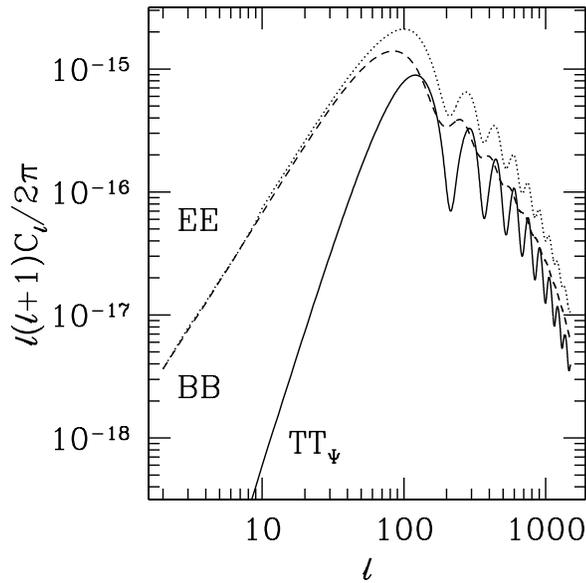


Figure 2.8: Contribution to the power spectra due to the source term $g\Psi$. The three curves show the power spectra calculated from Eqs. (2.21), (2.22), and (2.23). In the case of the temperature power spectrum the ISW sourcing term, $-\dot{h}e^{-\kappa}$, was neglected. Temperature (solid curve), E mode (dotted curve), and B mode (dashed curve). As all three curves are calculated from the same source term the differences are solely due to the effects of the differing projection functions.

extended nature of the source term.

The complicated form of the projection factors makes analytic progress difficult. Similarly, their oscillatory behaviour makes numerical integration very slow at high k values. CMBFAST implements a scheme for fast numerical integration. Here we discuss time averaging the projection functions to get a useful analytic envelope.

Widely known approximations for the spherical Bessel functions in the cases $x \gg l$ and $x \ll l$ exist and are in common usage. For our purposes, though, we are most interested in the case where $x \approx l$; i.e., in the vicinity of the peaks of the projection factor. A relatively simple approximate form may be derived which is valid in the regime $x > l$ (Debye, 1909),

$$j_l(x) = \frac{1}{\sqrt{x^2 \sin \alpha}} \cos [x(\sin \alpha - \alpha \cos \alpha) - \pi/4], \quad (2.47)$$

where $\cos \alpha = (l + 1/2)/x$. This can be shown to reduce to the usual $j_l(x) \approx \sin(x - l\pi/2)/x$ for $x \gg l$. Approximations valid in the regime $x \approx l$ exist, but are more complicated and will not significantly improve on this level of approximation.

This approximation is still complicated and shows strong oscillation. When we calculate power spectra we will be interested in quantities of the form $[P_X(x)]^2$. To proceed, we substitute Eq. (2.47) into the projection factor and then average the squared projection function over a full cycle to extract the variation of the envelope. Time averaging makes use of the relations $\langle \sin^2 x \rangle = \langle \cos^2 x \rangle = 1/2$ and $\langle \sin x \cos x \rangle = 0$. This envelope may then be further simplified by assuming $x \gg 1$ and $l \gg 1$. The resulting envelope functions are not pretty, but may be numerically integrated by a standard routine. They are

$$\langle P_T(x)^2 \rangle \approx \frac{1}{2x^5 \sqrt{x^2 - l^2}}, \quad (2.48)$$

$$\begin{aligned} \langle P_E(x)^2 \rangle \approx & \{ -16(l + l^2 - x^2)(l + 12l^3 + 8l^4 + 8x^4 - 4(x + 2lx)^2)^2 \\ & + (-16l^5(2 + l) - 4(1 + 2l)(3 + 10l)x^4 + 32x^6 \\ & + (-1 + 8l(1 + 2l))(x + 2lx)^2)^2 \} / \{ 512x^5(-l(1 + l) + x^2)^{9/2} \}, \quad (2.49) \end{aligned}$$

$$\langle P_B(x)^2 \rangle \approx \frac{(12l^2 - 8x^2)^2 - 16(l + 2l^2 - 2x^2)^2(l + l^2 - x^2)}{x^3(-4l^2 + 4x^2)^{5/2}}. \quad (2.50)$$

The form for $\langle P_T(x)^2 \rangle$ is consistent with that quoted by Zaldarriaga & Harari (1995).

One slight complication is that these approximations show divergent behaviour as $x \rightarrow l$ making it necessary to arbitrarily restrict the domain to $x > l + a$, where a is an arbitrary cutoff of order unity. Fig. 2.7 shows these approximations and the cutoff. This cutoff procedure behaves best in the case of the B projection factor which is already decreasing as $x \rightarrow l$. The E and T projection factors have considerable weight near to the peak, and so care must be taken in selecting the cutoff.

To apply these projection factors, we approximate the anisotropy term,

$$\Delta_{Xl} = \int_0^{\tau_0} d\tau g(\tau) \Psi(\tau) P_{Xl}[k(\tau_0 - \tau)], \quad (2.51)$$

by the expression,

$$\Delta_{Xl} \approx P_{Xl}[k(\tau_0 - \tau_R)] \int_0^{\tau_0} d\tau g(\tau) \Psi(\tau), \quad (2.52)$$

where we have pulled the projection factor out from the integral. This should be a good approximation provided that the projection factor varies slowly relative to the source term.

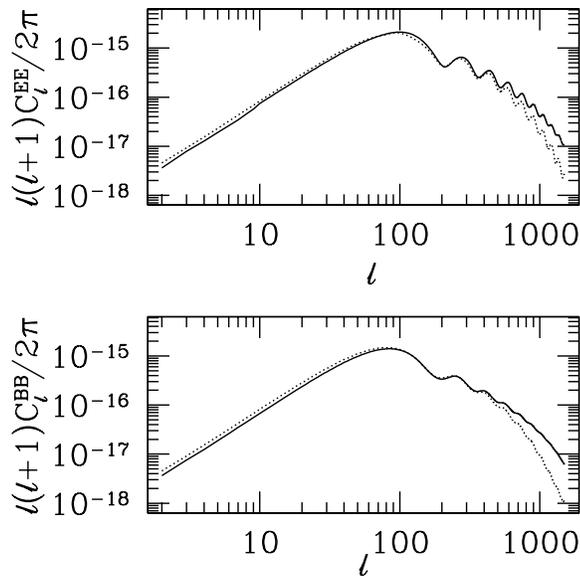


Figure 2.9: Comparison of power spectra with exact (solid curve) and approximate (dashed curve) projection factors. The approximate power spectra are calculated using Eq. (2.52) with the projection factors from Eqs. (2.49) and (2.50), and using source terms calculated by CMBFAST.

The power spectra calculated under this approximation are shown in Fig. 2.9. Note

that two approximations are combined here. The projection term has been pulled from the integral and then the time averaged expression used in place of the exact form. The inadequacy of this approximation at large l can be seen. This is to be expected, as in this regime the relevant k modes oscillate significantly over the width of the last-scattering surface. At values $l < 500$, the approximation is appropriate.

2.7 Source evolution

Having discussed the recombination history and the mechanics of projection, we turn to the core issue of how the source itself evolves.

Those modes that enter the horizon close to recombination evolve only slowly over the width of the last-scattering surface. For these modes, it is possible to derive analytic approximations to the source functions that occur in the expressions for $\tilde{\Delta}_X$. Here, we follow the approach of Zaldarriaga & Harari (1995), but see also Keating et al. (1998). Preceding recombination, the optical depth κ is large and the photons are tightly coupled to the baryonic fluid. In this regime, we may expand the Boltzmann equations for the temperature and polarisation multipoles in powers of $\dot{\kappa}^{-1}$. Keeping terms to first order in $\dot{\kappa}^{-1}$, we obtain the equations,

$$\dot{\tilde{\Delta}}_{T0} = -\dot{h} - \dot{\kappa}[\tilde{\Delta}_{T0} - \Psi], \quad (2.53)$$

$$\dot{\tilde{\Delta}}_{P0} = -\dot{\kappa}[\tilde{\Delta}_{P0} + \Psi], \quad (2.54)$$

$$\dot{\tilde{\Delta}}_{Tl} = 0, \quad l \geq 1, \quad (2.55)$$

$$\dot{\tilde{\Delta}}_{Pl} = 0, \quad l \geq 1. \quad (2.56)$$

Using these equations together with the definition of Ψ gives us an expression for the time evolution of the source function within this tightly-coupled limit,

$$\dot{\Psi} + \frac{3}{10}\dot{\kappa}\Psi = -\frac{\dot{h}}{10}. \quad (2.57)$$

Thus, an approximate solution for Ψ is

$$\Psi(\tau) = \int_0^{\tau_0} d\tau' \left(-\frac{\dot{h}(\tau')}{10} \exp \left[-\frac{3}{10} \left(\kappa(\tau') - \kappa(\tau) \right) \right] \right). \quad (2.58)$$

If we assume that the visibility function is approximately Gaussian during recombination, then we may approximate $\dot{\kappa} \approx -\kappa/\Delta\tau_R$. This allows a change of variable to $x = \kappa(\tau')/\kappa(\tau)$ which leads to

$$\Psi(\tau) = -\frac{\dot{h}(\tau_R)}{10} e^{\frac{3}{10}\kappa(\tau)} \Delta\tau_R \int_1^\infty \frac{dx}{x} e^{-\frac{3}{10}\kappa x}. \quad (2.59)$$

In taking the gravitational-wave driving term outside of the integral, we have assumed that h varies slowly over the visibility function. This approximation is only valid for $k \ll 1/\Delta\tau_R$. At larger wavenumbers, the rapid oscillation of Ψ over the visibility function makes this a poor approximation. An improvement is to replace $\dot{h}(\tau_R)$ in Eq. (2.59) with its value averaged over the visibility function

$$\langle \dot{h}(\tau) \rangle = \int_0^{\tau_0} d\tau g(\tau) \dot{h}(\tau) \approx \dot{h}(\tau_R) e^{-(k\Delta\tau_R)^2/2}. \quad (2.60)$$

In calculating the right-hand side, we have treated \dot{h} as an oscillatory function with a slowly-varying envelope. Integrating an oscillatory function over a Gaussian leads to the function evaluated at the Gaussian's peak multiplied by a decaying exponential. This exponential decay has a clear physical interpretation. For modes with $k > 1/\Delta\tau_R$, the source function oscillates rapidly across the visibility function. Hence, different regions in the SLS contribute to the observed polarisation with different phases leading to cancellation and a decrease in the observed power. We will refer to this cancellation as phase-damping. While present in the scalar modes, this effect is overwhelmed by diffusion damping (Hu & Sugiyama, 1996). Diffusion damping makes the effective visibility function scale dependent and always sufficiently narrow that phase-damping is not important. For the tensor modes, phase-damping provides the dominant process for damping on small scales.

Fig. 2.10 shows the behaviour of Ψ and our analytic approximations for four values of k . For small k , the approximation closely mirrors the growth of Ψ in the region where the visibility function has weight. At larger k , the source function is seen to oscillate across the width of $g(\tau)$; this is not reproduced by either approximation. This should not be cause for concern as we now discuss.

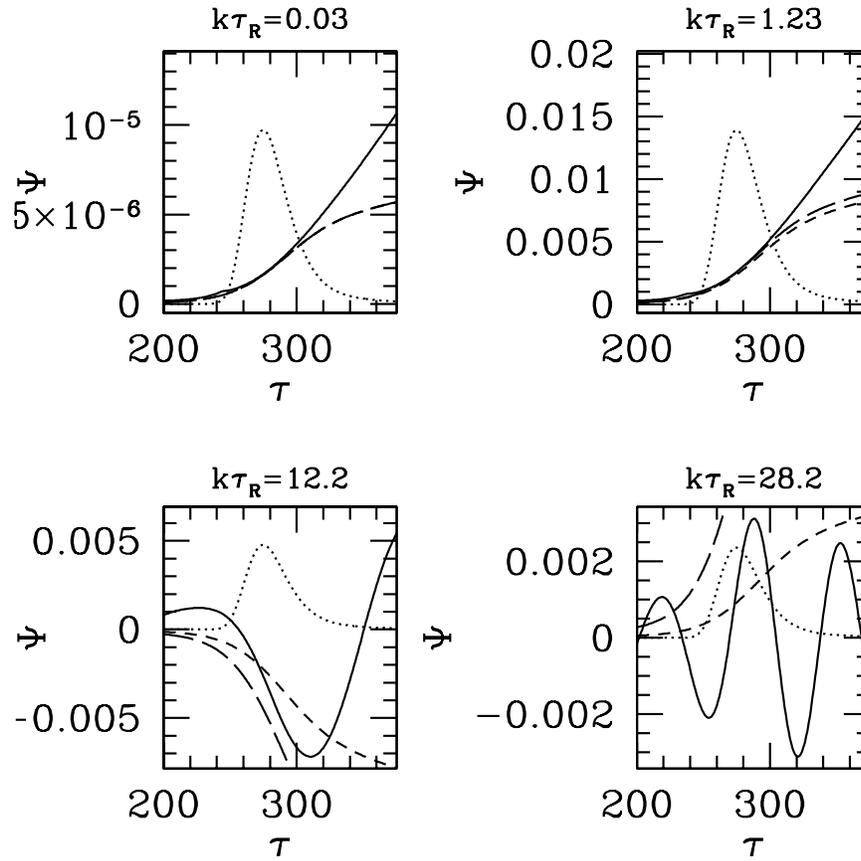


Figure 2.10: Comparison of Eq. (2.59) (long dashed curve), Eq. (2.59) with damping term (short dashed curve), and numerical calculation of Ψ (solid curve). Four different values of k are plotted: $k\tau_R = 0.03, 1.23, 12.2,$ and 28.2 . An arbitrarily scaled visibility function (dotted curve) has been plotted in each panel to guide the eye.

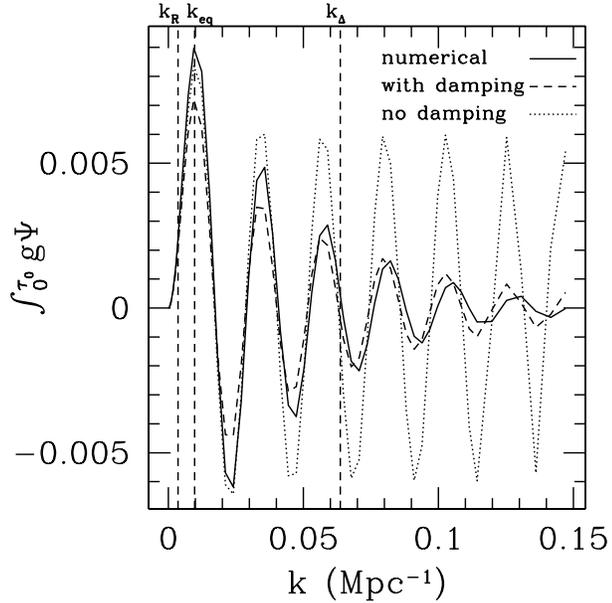


Figure 2.11: Comparison of $\int_0^{\tau_0} g\Psi$ for the numerical Ψ (solid curve), the approximation for Ψ without phase-damping (dotted curve) and the approximation for Ψ with phase-damping (dashed curve). Vertical lines indicate (from left to right) $1/\tau_R$, $1/\tau_{\text{eq}}$, and $1/\Delta\tau_R$.

The quantity of real interest is the anisotropy that this source generates. This is calculated from Eqs. (2.22) and (2.23). For the given polarisation, $\mathbf{X}=(\text{E},\text{B})$, we have

$$\Delta_{Xl}(k) = \int_0^{\tau_0} d\tau g(\tau)\Psi(\tau)P_{Xl}[k(\tau_0 - \tau)]. \quad (2.61)$$

First we pull the projection term outside of the integral assuming that it varies slowly over the width of the visibility function,

$$\Delta_{Xl}(k) \approx P_{Xl}[k(\tau_0 - \tau_R)] \int_0^{\tau_0} d\tau g(\tau)\Psi(\tau). \quad (2.62)$$

Note that Ψ appears only through an integral over the visibility function. Provided that our approximation can reproduce this integrated behaviour, the fact that it fails to reproduce the temporal oscillation is unimportant. Fig. 2.11 indicates the close agreement between the integrated source function and our approximation, provided that phase-damping is taken into account.

Having checked the validity of our approximation we substitute for Ψ in (2.62) using

Eqs. (2.59) and (2.60) giving

$$\Delta_{Xl}(k) = P_{Xl}[k(\tau_0 - \tau_R)] \frac{1}{10} \dot{h}(\tau_R) \Delta\tau_R e^{-(k\Delta\tau_R)^2/2} \int_0^\infty d\kappa e^{-\frac{7}{10}\kappa} \int_1^\infty \frac{dx}{x} e^{-\frac{3}{10}\kappa x}. \quad (2.63)$$

The integrals evaluate to $(10/7) \log(10/7)$, which leads to the final result

$$\Delta_{Xl} = P_{Xl}[k(\tau_0 - \tau_R)] \dot{h}(\tau_R) \Delta\tau_R e^{-(k\Delta\tau_R)^2/2} \left(\frac{1}{7} \log \frac{10}{3} \right). \quad (2.64)$$

This result is proportional to the width $\Delta\tau_R$ of recombination as might be expected. During recombination, photons will travel for a distance of order $\Delta\tau_R$ before scattering. This is the time available for the quadrupole which sources the polarisation to grow, and so we expect a result proportional to $k\Delta\tau_R$.

Extending this result to calculate the power spectrum is straightforward. We have Eq. (2.26),

$$C_{Xl} = (4\pi)^2 \int k^2 dk P_h(k) [\Delta_{Xl}(k)]^2. \quad (2.65)$$

Applying our expression for $\Delta_{Xl}(k)$ yields the final result for this Section,

$$C_{Xl} = (4\pi)^2 \left(\frac{1}{7} \log \frac{10}{3} \right)^2 \int k^2 dk P_h(k) P_{Xl}[k(\tau_0 - \tau_R)]^2 \dot{h}(\tau_R)^2 \Delta\tau_R^2 e^{-(k\Delta\tau_R)^2}. \quad (2.66)$$

Fig. 2.12 compares the result of CMBFAST with that from Eq. (2.66). All plots have been calculated for the fiducial cosmology with $\Omega_b = 0.05$, $\Omega_{DM} = 0.25$, and $\Omega_\Lambda = 0.7$. Values for τ_R and $\Delta\tau_R$ were chosen by fitting by eye to the visibility function produced by CMBFAST. Radiation and numerical forms for h without phase-damping are plotted alongside a numerical form for h with phase-damping and the results of the full numerical calculation.

Agreement between all of the solutions is good at low l where our assumptions are most valid and the anisotropy is building slowly. Similarly, the position of the main peak is accurately reproduced, although it becomes clear that the E projection-factor approximation is less reliable than the B-mode one. Beyond the main peak, the effects of the different forms for h become apparent.

Without phase-damping, the radiation-dominated form of h leads to an almost flat power spectrum, consistent with Eq. (2.41), with a pattern of bumps and wiggles roughly

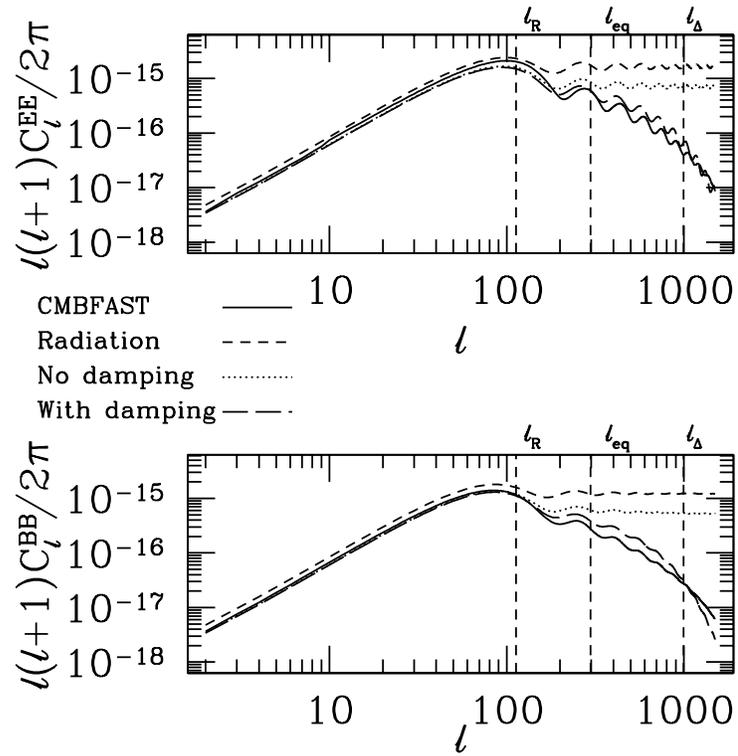


Figure 2.12: Comparison of E and B power spectra for CMBFAST (solid curve) and analytic approximations: h for radiation epoch (short dashed curve), h calculated numerically (dotted curve), and h calculated numerically with exponential damping (long dashed curve).

in phase with those in the full numerical calculation. Moving to the numerical form of h leads to some damping over the radiation case. This reflects the increased redshifting which occurs when matter becomes important between τ_{eq} and τ_R . We never observe the expected matter-dominated scaling of l^{-2} , although the presence of matter does lead to a slight decline in power as l increases. For this cosmology, the ratio ρ_r/ρ_m evaluated at τ_R is 0.29 showing that the radiation content is still significant at recombination.

Neither of these forms reproduces the rapid decline in power at large l , which is not surprising as we have yet to include phase-damping. Once this is included, the shape of the power spectrum is much closer to that of the numerical calculation showing a sharp decline in power above $l \approx 300$ and reproducing the position of the peaks to reasonable accuracy. Power in the range $l = 150$ to $l = 600$ is slightly overestimated. For $l > 600$, the limitations of our projection-factor approximations become apparent with the very rapid drop in power previously observed in Section 2.6.

2.8 Discussion

Here, we discuss the information that the features in the tensor power spectrum contain and how detection would complement our existing understanding of the early Universe. In contrast to the scalar modes, the tensor modes contain very clean information about the evolution of the Universe. The features of the scalar spectra are a result of the oscillation of the matter-radiation fluid during the period up to recombination. The scalar spectra encode information about the sound speed of the baryon-radiation fluid, the baryon fraction, and other information about the particle content of the Universe (Dodelson, 2003; Hu et al., 2001). In contrast, the features of the tensor spectrum are determined solely by the wave motion of the evolving gravitational waves. They primarily contain information about the expansion rate during the early Universe. Through their overall angular scale, both spectra encode basic information about the epoch and duration of recombination and the geometry of the Universe.

The first peak of the tensor polarisation spectrum, occurring at $l \approx 90$ for the B modes and $l \approx 105$ for the E modes, is determined by the horizon scale at recombination. The exact angular scale is determined by this along with the redshift of recombination and the geometry of the Universe. This information can be determined from the scalar modes

allowing a direct measurement of the horizon scale. The amplitude of the tensor power spectra is directly related to the energy scale of inflation. Slow-roll inflation, parameterised by the energy scale E_i of inflation, predicts a B-mode power spectrum with a peak at $l \approx 90$ and peak amplitude (Hu et al., 2003)

$$\Delta B_{\text{peak}} = 0.024 \left(\frac{E_i}{10^{16} \text{GeV}} \right)^2 \mu\text{K}. \quad (2.67)$$

Measuring this main peak is the subject of several experimental endeavours (BICEP, QUIET, POLARBEAR) with hope of detection in the not-so-distant future.

For all models consistent with WMAP constraints to the energy scale of inflation, features after the main peak are sub-dominant to the lensed B-mode signal (Hu et al., 2003; Wang et al., 2002). This necessitates the use of algorithms to clean the polarisation maps and recover the tensor signal. Techniques using maximum likelihood (Seljak & Zaldarriaga, 1999; Hirata & Seljak, 2003a,b) and quadratic estimators (Hu & Okamoto, 2002; Kesden et al., 2003) have been advanced to deal with this problem. Even so, this will complicate precision measurements of the tensor B mode after the main peak.

Measuring the overall amplitude after the main peak should recover the scaling relations discussed in Section 2.4. The breaks in the different regimes yield the horizon scales of matter-radiation equality and the width of the SLS. This in itself is enough information to constrain a cosmological model yielding Ω_m and Ω_r . Neutrino anisotropic stress further damps power on small scales and will make detection more difficult while adding extra information about the neutrino fraction.

The positions of the peaks and troughs on small scales contain information about the phase of the gravitational wave at recombination. This in turn depends upon the early expansion rate. The acoustic peaks in the scalar power spectrum can be used as a standard ruler; the wiggles in the tensor power spectrum can have the same utility, but operating on a different range and spacing. Measuring these wiggles would better constrain our cosmology, though it is doubtful that scales small enough to probe the very early Universe will be observed in the foreseeable future.

2.9 Conclusions

This investigation has probed the individual elements that compose the calculation of the tensor power spectra. Using a variety of approximations, we have obtained a semi-analytic expression which qualitatively reproduces the behaviour of more detailed calculation. While the approximation is clearly not suitable for precise comparison with data, it serves to illustrate the important physics in an intuitive fashion. We have shown that the features of the power spectrum may be explained with reference to three main scales: the horizon size at recombination, the horizon size at matter-radiation equality, and the width of the SLS. The first two scales determine the evolution of the tensor modes; the latter relates to the effect of the thermal history on the generation of anisotropies. The shapes of the polarisation spectra show most sensitivity to the width of the SLS through the phenomenon of phase-damping that dominates on small scales. The effect of the thermal history on the temperature spectra is much less dramatic, affecting the amplitude and smoothing on small scales. We have seen that the position of the peaks and troughs in the power spectrum relate to the phase of the gravitational wave at recombination. It is of interest that we do not see modes displaying matter-dominated behaviour in our calculations. This would allow the tensor spectra to probe the radiation content at recombination. Useful scaling relations have been developed and clarified. We hope this paper will aid in general understanding of the tensor modes and inspire future experimental efforts.

Acknowledgments

This work was supported in part by NASA NAG5-11985, and DoE DE-FG03-92ER40701.

Chapter 3

Descending from on high: Lyman series cascades and spin-kinetic temperature coupling in the 21 cm line

We examine the effect of Lyman continuum photons on the 21 cm background in the high-redshift universe. The brightness temperature of this transition is determined by the spin temperature T_s , which describes the relative populations of the singlet and triplet hyperfine states. Once the first luminous sources appear, T_s is set by the Wouthuysen-Field effect, in which Lyman-series photons mix the hyperfine levels. Here we consider coupling through $n > 2$ Lyman photons. We first show that coupling (and heating) from scattering of $\text{Ly}n$ photons is negligible, because they rapidly cascade to lower-energy photons. These cascades can result in either a $\text{Ly}\alpha$ photon – which will then affect T_s according to the usual Wouthuysen-Field mechanism – or photons from the $2s \rightarrow 1s$ continuum, which escape without scattering. We show that a proper treatment of the cascades delays the onset of strong Wouthuysen-Field coupling and affects the power spectrum of brightness fluctuations when the overall coupling is still relatively weak (i.e., around the time of the first stars). Cascades damp fluctuations on small scales because only $\sim 1/3$ of $\text{Ly}n$ photons cascade through $\text{Ly}\alpha$, but they do not affect the large-scale power because that arises from those photons that redshift directly into the $\text{Ly}\alpha$ transition. We also comment on the utility of $\text{Ly}n$ transitions in providing “standard rulers” with which to study the high-redshift universe.

3.1 Introduction

One potentially promising probe of the cosmic dark ages is 21 cm tomography. It has long been known (Hogan & Rees, 1979; Scott & Rees, 1990) that neutral hydrogen in the intergalactic medium (IGM) may be detectable in emission or absorption against the cosmic microwave background (CMB) at the wavelength of the redshifted 21 cm line, the spin-flip transition between the singlet and triplet hyperfine levels of the hydrogen ground state. The brightness of this transition will thus trace the distribution of HI in the high-redshift universe (Field, 1958, 1959a), which gives the signal angular structure as well as structure in redshift space. These features arise from inhomogeneities in the gas density field, the hydrogen ionization fraction and the spin temperature. Madau, Meiksin & Rees (1997) showed that the first stars could cause a rapid evolution in the signal through their effect on the spin temperature. Consequently, the 21 cm signal can provide unparalleled information about the “twilight zone” when the first luminous sources formed and the epoch of reionization and reheating commenced.

Despite the theoretical promise of this probe, it is only with improvements in computing power that building radio arrays with sufficient sensitivity, capable of correlating billions of visibility measurements, has become possible (Morales & Hewitt, 2004). Three such arrays (LOFAR¹, MWA², and PAST³) will soon be operational, opening a window onto this new low frequency band. Before a detection can be made, however, there are still major scientific and technical challenges to be met. Ionospheric scattering and terrestrial interference are two serious issues. Also worrying is the need to remove foregrounds, which are many orders of magnitude stronger than the signal. Multifrequency subtraction techniques (Zaldarriaga, Furlanetto & Hernquist, 2004; Morales & Hewitt, 2004; Santos, Cooray & Knox, 2005), exploiting the smoothness of the foreground spectra, have been proposed, but their effectiveness has yet to be tested. The challenges are great, but so are the opportunities. It is thus crucial to understand the nature of the 21 cm signal as we commence these searches.

Fluctuations in the 21 cm signal arise from both cosmological and astrophysical sources. Most previous work has focussed on the signal due to density perturbations (Madau et al., 1997; Loeb & Zaldarriaga, 2004) or from inhomogeneous ionization (Ciardi & Madau, 2003;

¹See <http://www.lofar.org/>.

²See <http://web.haystack.mit.edu/arrays/MWA/>.

³See Pen et al. (2005).

Furlanetto, Sokasian & Hernquist, 2004; Furlanetto, Zaldarriaga & Hernquist, 2004). An additional source of fluctuations is the spin temperature, which describes the relative occupation of the singlet and triplet hyperfine levels. These levels may be excited by three primary mechanisms: absorption of CMB photons, atomic collisions, and absorption and re-emission of Ly α photons (the Wouthuysen-Field effect; Wouthuysen, 1952; Field, 1959a). The first two processes rely upon simple physics, but the last one allows us to study the properties of luminous sources, which determine the background radiation field.

Barkana & Loeb (2005b, henceforth BL05) studied the signal generated by the first generation of collapsed objects. These high redshift objects are highly biased, leading to large variations in their number density. This, combined with the $1/r^2$ dependence of the flux, causes large fluctuations in the Ly α background, which can be probed through their effect on the 21 cm transition. Exploiting the anisotropy induced by peculiar velocities (Bharadwaj & Ali, 2004; Barkana & Loeb, 2005a), they showed that information about the Ly α radiation field could be extracted from the power spectrum of 21 cm fluctuations and separated into those fluctuations correlated and uncorrelated with the density field. The features of these spectra allow extraction of astrophysical parameters such as the star formation rate and bias. However, it is not a trivial task to relate the emissivity to a distribution of Ly α photons. The background in the Ly α line is composed of two parts: those photons that have redshifted directly to the Ly α frequency and those produced by atomic cascades from higher Lyman series photons. To calculate this latter component, BL05 assumed that atomic cascades were 100% efficient at converting photons absorbed at a Lyman resonance into a Ly α photon, while in reality most cascades end in two-photon decay from the $2S$ level.

In this paper, we calculate the exact cascade conversion probabilities from basic atomic physics. In addition, we discuss the possibility of level mixing by scattering of Ly n photons via a straightforward generalisation of the Wouthuysen-Field effect. We then apply the cascade efficiencies to calculate the Ly α flux profile of an isolated source. The existence of discrete horizons, determined by the maximum distance a photon can travel before it redshifts into a given Lyman resonance, imprints a series of discontinuities into the profile, which can in principle be used as a standard ruler. We apply these results to the power spectra of 21 cm fluctuations during the epoch of the first stars, showing that these corrections cannot be ignored when extracting astrophysical parameters.

The layout of this paper is as follows. In §3.2 we introduce the formalism for describing 21 cm fluctuations and the dominant coupling mechanism, the Wouthuysen-Field effect. In §3.3 we discuss the possibility of direct pumping by Ly α photons. Next, in §3.4, we detail the atomic physics of radiative cascades in atomic hydrogen. The results are applied to the Ly α flux profile of an isolated source in §3.5 and to the 21 cm power spectrum from the first galaxies in §3.6. We also discuss some of the limitations of this formalism. Finally, we summarise our results in §3.7. In an Appendix, we review the equations needed to calculate analytically the Einstein A coefficients for the hydrogen atom. Throughout, we assume $(\Omega_m, \Omega_b, \Omega_\Lambda, h, \sigma_8, n_s) = (0.3, 0.046, 0.7, 0.7, 0.9, 1.0)$, consistent with the most recent measurements (Spergel et al., 2003).

During the preparation of this paper, Hirata (2006) submitted a preprint covering similar material. We have confirmed agreement where there is overlap. The main results of this work were discussed at the “Reionizing the Universe” conference in Groningen, The Netherlands (June 27-July 1, 2005; see <http://www.astro.rug.nl/~cosmo05/program.html>).

3.2 21 cm formalism and the Wouthuysen-Field mechanism

The 21 cm line of the hydrogen atom results from hyperfine splitting of the $1S$ ground state due to the interaction of the magnetic moments of the proton and the electron. The HI spin temperature T_s is defined via the relative number density of hydrogen atoms in the $1S$ singlet and triplet levels $n_1/n_0 = (g_1/g_0) \exp(-T_\star/T_s)$, where $(g_1/g_0) = 3$ is the ratio of the spin degeneracy factors of the two levels, and $T_\star \equiv hc/k\lambda_{21\text{cm}} = 0.0628$ K. The optical depth of this transition is small at all relevant redshifts, so the brightness temperature of the CMB is

$$T_b = \tau \left(\frac{T_s - T_{\text{CMB}}}{1 + z} \right), \quad (3.1)$$

where the optical depth for resonant 21 cm absorption is

$$\tau = \frac{3c\lambda^2 h A_{10} n_{\text{HI}}}{32\pi k_B T_s (1 + z) (dv_r/dr)}. \quad (3.2)$$

Here n_{HI} is the number density of neutral hydrogen, $A_{10} = 2.85 \times 10^{-15} \text{ s}^{-1}$ is the spontaneous emission coefficient, and dv_r/dr is the gradient of the physical velocity along the line of sight with r the comoving distance. When $T_s < T_{\text{CMB}}$ there is a net absorption of CMB

photons, and we observe a decrement in the brightness temperature.

The spin temperature is determined by three coupling mechanisms. Radiative transitions due to absorption of CMB photons (as well as stimulated emission) tend to drive $T_s \rightarrow T_{\text{CMB}}$. Spin flips from atomic collisions drive $T_s \rightarrow T_k$, the gas kinetic temperature. Finally, the Wouthuysen-Field effect (Wouthuysen, 1952; Field, 1958), which is the main focus of this paper, also drives $T_s \rightarrow T_k$ (see below). The combination that appears in (3.1) can be written

$$\frac{T_s - T_{\text{CMB}}}{T_s} = \frac{x_{\text{tot}}}{1 + x_{\text{tot}}} \left(1 - \frac{T_{\text{CMB}}}{T_k} \right), \quad (3.3)$$

where $x_{\text{tot}} = x_\alpha + x_c$ is the sum of the radiative and collisional coupling parameters. The latter is

$$x_c = \frac{4\kappa_{1-0}(T_k)n_H T_\star}{3A_{10}T_{\text{CMB}}}, \quad (3.4)$$

where κ_{1-0} is tabulated as a function of T_k (Allison & Dalgarno, 1969; Zygelman, 2005). The spin temperature becomes strongly coupled to the gas temperature when $x_{\text{tot}} \gtrsim 1$.

A schematic diagram of the Wouthuysen-Field effect is shown in Figure 3.1; it mixes the hyperfine levels through absorption and re-emission of Ly α photons. Quantum selection rules allow transitions for which the total spin angular momentum F changes by $\Delta F = 0, \pm 1$ (except $0 \rightarrow 0$), making only two of the four $n = 2$ levels accessible to both the $n = 1$ singlet and triplet states. Transitions to either of these states can change T_s . The coupling coefficient is

$$x_\alpha = \frac{4P_\alpha T_\star}{27A_{10}T_{\text{CMB}}}, \quad (3.5)$$

where P_α is the Ly α scattering rate (Madau et al., 1997). If resonant scattering of Ly α photons occurs rapidly enough T_s will be driven to T_α , the colour temperature of the radiation field at the Ly α frequency (Field, 1958; Madau et al., 1997). In parallel, the repeated scattering of Ly α photons by the thermal distribution of atoms brings $T_\alpha \rightarrow T_k$ (Field, 1959b; Hirata, 2006). Consequently, the Wouthuysen-Field effect provides an effective coupling between the spin temperature and the gas kinetic temperature.

We can also write the Wouthuysen-Field coupling as

$$x_\alpha = \frac{16\pi^2 T_\star e^2 f_\alpha}{27A_{10}T_{\text{CMB}}m_e c} S_\alpha J_\alpha, \quad (3.6)$$

where $f_\alpha = 0.4162$ is the oscillator strength for the Ly α transition, S_α is a correction

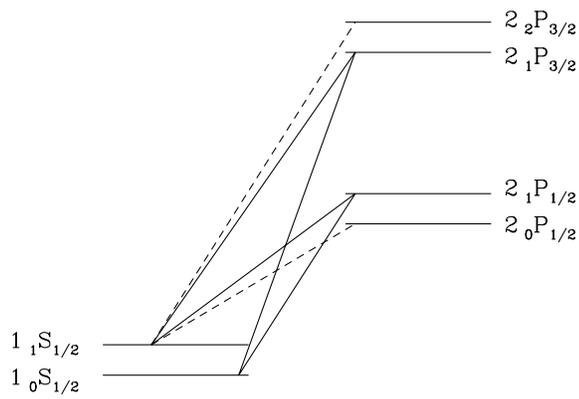


Figure 3.1: Hyperfine structure of the $2P$ and $1S$ level of the hydrogen atom. Levels are labelled according to the notation $n_F L_J$, where n , L , and J are the usual radial, orbital angular momentum and total angular momentum quantum numbers. $F = I + J$ is the quantum number obtained from the nuclear spin and J . Allowed transitions obey $\Delta F = 0, \pm 1$ (except $0 \rightarrow 0$). Those relevant for the Wouthuysen-Field effect are indicated by solid curves, while dashed curves indicate the remaining allowed transitions.

factor of order unity (Chen & Miralda-Escudé, 2004; Hirata, 2006) that accounts for the redistribution of photon energies due to repeated scattering off the thermal distribution of atoms, and J_α is the angle-averaged specific intensity of Ly α photons by photon number. For reference, a Ly α flux of $J_\alpha = 1.165 \times 10^{-10} [(1+z)/20] \text{ cm}^{-2} \text{ s}^{-1} \text{ Hz}^{-1} \text{ sr}^{-1}$ yields $x_\alpha = 1$ (corresponding to $P_\alpha = 7.85 \times 10^{-13} [1+z] \text{ s}^{-1}$).

Fluctuations in the brightness temperature arise from fluctuations in the density, the Wouthuysen-Field coupling, the neutral fraction x_{HI} and the radial velocity component. To linear order

$$\delta_{T_b} = \beta\delta + \frac{x_\alpha}{\tilde{x}_{\text{tot}}}\delta_{x_\alpha} + \delta_{x_{\text{HI}}} - \delta_{d_r v_r}, \quad (3.7)$$

where δ_a is the fractional perturbation in a , δ is the fractional density perturbation, and $\tilde{x}_{\text{tot}} = x_{\text{tot}}(1 + x_{\text{tot}})$. β is a parameter describing the thermal history of the gas, which we assume to have cooled adiabatically, so that $\beta \approx 0.2$. Naoz & Barkana (2005) showed that β slightly increases and exhibits mild scale dependence, as gas temperature fluctuations do not exactly track the density fluctuations. We note that, on scales $0.01 \text{ Mpc}^{-1} < k < 10^3 \text{ Mpc}^{-1}$, β is approximately constant at $z = 20$ and choose to ignore this subtlety for ease of comparison with BL05. The first three components of equation (3.7) are isotropic, but the velocity fluctuation introduces an anisotropy of the form $\delta_{d_r v_r}(k) = -\mu^2\delta$ in Fourier space (Bharadwaj & Ali, 2004), where μ is the cosine of the angle between the wavenumber \mathbf{k} of the Fourier mode and the line of sight. This allows us to separate the brightness temperature power spectrum P_{T_b} into powers of μ^2 (Barkana & Loeb, 2005a)

$$P_{T_b}(\mathbf{k}) = \mu^4 P_{\mu^4}(\mathbf{k}) + \mu^2 P_{\mu^2}(\mathbf{k}) + P_{\mu^0}(\mathbf{k}). \quad (3.8)$$

The anisotropy is sourced only by density fluctuations, so that P_{μ^4} depends only on the matter power spectrum. P_{μ^2} contains cross-correlations between matter fluctuations and both δ_{x_α} and $\delta_{x_{\text{HI}}}$, making it an ideal probe of fluctuations in the radiation background. In particular, at sufficiently high redshifts such that $x_{\text{HI}} \ll 1$, it probes variations in the Ly α background. Linear combinations of these three terms can be used to extract detailed information about other types of fluctuations (Barkana & Loeb, 2005a).

3.3 Direct pumping by Lyman series photons

Of course, the radiation background contains photons that redshift into all the Lyman transitions, not just Ly α . The main purpose of this paper is to examine how these affect T_s . The existing literature assumes that all Ly n photons are immediately converted into Ly α photons by atomic cascades (e.g. BL05). In reality, there are two different contributions to consider: one due to scattering of the Ly n photon itself and the other due to its cascade products. In this section, we discuss the direct contribution of Ly n scattering to the coupling of T_s and T_k , which occurs in a manner exactly analogous to the Wouthuysen-Field effect. For this effect to be significant two requirements must be fulfilled. First, the scattering rate of Ly n photons must be sufficient to couple T_s and T_n , the Ly n colour temperature. Second, it must be sufficient to drive $T_n \rightarrow T_k$. We will argue that neither condition is satisfied in practice.

The IGM is optically thick $\tau \gg 1$ to all Lyman series transitions with $n \lesssim 100$. Consequently, a Ly α photon emitted by a star will scatter many times ($\sim \tau \sim 10^6$; see Gunn & Peterson, 1965) before it finally escapes by redshifting across the line width; each of these scatterings contributes to the Wouthuysen-Field coupling. A Ly n photon can escape by redshifting across the line width, but a transition to a level other than $n = 1$ will also remove it. The probability for a decay from an initial state i to a final state f is given in terms of the Einstein A_{if} coefficients by

$$P_{if} = \frac{A_{if}}{\sum_f A_{if}}. \quad (3.9)$$

Appendix B summarises the expressions needed to compute the Einstein A_{if} coefficients. For the Lyman series transitions $P_{nP \rightarrow 1S} \approx 0.8$ (see Table 1) so that a Ly n photon will scatter of order $N_{\text{scat}} \approx 1/(1 - P_{nP \rightarrow 1S}) \sim 5$ times before undergoing a cascade.

Because a cascade occurs long before escape via redshifting, the coupling from direct pumping is negligible. Recall that the scattering rate P_X for the photon type $X = \text{Ly}\alpha$, Ly β , etc. may be expressed as (Field, 1959a)

$$n_{\text{HI}} P_X = N_{\text{scat}} \dot{n}_X \quad (3.10)$$

in terms of the production rate of photons per unit volume \dot{n}_X . It is then clear that,

for similar production rates (i.e., for sources with a reasonably flat spectrum), $P_n/P_\alpha \sim N_{\text{scat},n}/N_{\text{scat},\alpha} \sim 5 \times 10^{-6}$. This simple argument shows that the contribution from direct pumping by Ly n photons will be negligible compared to that of the Ly α photons, because $x_\alpha \propto P_\alpha$.

The second question, whether $T_n \rightarrow T_k$, is still relevant for heating of the gas by repeated scatterings. Given the reduced number of scattering events, it seems unlikely to be the case, but a full calculation using a Monte Carlo method or following Chen & Miralda-Escudé (2004) is required to rigourously answer this question. Lack of equilibrium would make Ly n scattering a more efficient source of heat, on a per scattering basis. Chen & Miralda-Escudé (2004) have shown that Ly α heating is much smaller than previous calculations indicated (Madau et al., 1997), because $T_\alpha \approx T_k$, which reduces the heat transferred per collision. This is unlikely to be the case for the Ly n .

Following Madau et al. (1997), we can estimate the *maximum* heating from a single Ly n scattering by assuming that all of the atomic recoil energy for a stationary atom is deposited in the gas. Momentum conservation then demands

$$\dot{E}_n = - \left\langle \frac{\Delta E}{E} \right\rangle h\nu_n P_n, \quad (3.11)$$

where $\langle \Delta E/E \rangle \sim 10^{-8}$ is the fraction of energy lost by a Ly n photon after scattering from a stationary hydrogen atom, and $h\nu_n$ is the energy of the photon. Assuming the production rate of Ly n photons is comparable to that of the Ly α photons and taking $x_\alpha = 1$, we then obtain $\dot{E}_n \sim 0.002[(1+z)/10] \text{ K Gyr}^{-1}$. This is much smaller than the Ly α heating rate, even including the $T_\alpha \approx T_k$ correction, so we do not expect Ly n scattering to be a significant heat source. Furthermore, if $T_n \approx T_k$ the rate would be much smaller than this estimate, as in Chen & Miralda-Escudé (2004).

3.4 Lyman series cascades

An excited state of hydrogen may reach the ground state in three ways. Firstly, it may decay directly to the ground state from an nP state ($n > 2$), generating a Ly n photon. Secondly, it may cascade to the metastable $2S$ level. Decay from the $2S$ level proceeds via a forbidden two-photon process. Finally, it may cascade to the $2P$ level, from which it will

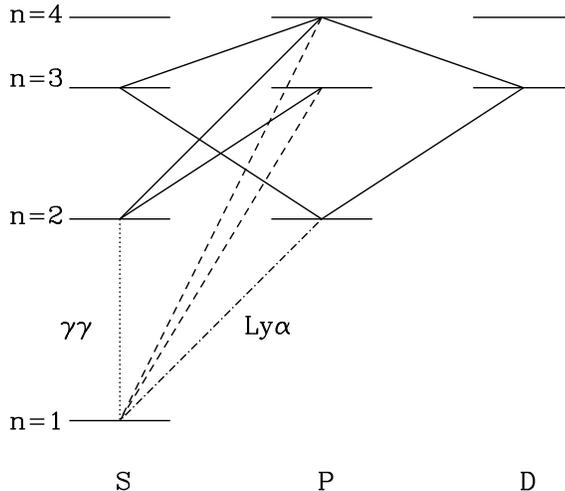


Figure 3.2: Energy level diagram for the hydrogen atom illustrating Ly β and Ly γ cascades. Marked decays are distinguished as cascades (solid curves), Ly n (dashed curves), Ly α (dot-dashed curve), and the two-photon decay (dotted curve). Note that the selection rules ($\Delta L = \pm 1$) decouple the 3P and 2P levels, preventing Ly β from being converted into Ly α .

produce a Ly α photon. We are primarily interested in the fraction of decays that generate Ly α photons, which will increase the Ly α flux pumping the hyperfine levels.

The fraction of cascades that generate Ly α photons can be determined straightforwardly from the selection rules and the decay rates. As an example, consider the Ly β system. Absorption of a Ly β photon excites the atom into the 3P level. As illustrated in Figure 3.2, the 3P level can decay directly to the ground state, regenerating the Ly β photon, or to the 2S level, where it will decay by two-photon emission. The selection rules forbid Ly β photons from being converted into Ly α photons. In contrast, the 4P level, excited by absorption of Ly γ , can cascade via the 3S or 3D levels to the 2P level and then generate Ly α .

To calculate the probability f_{recycle} that a Ly n photon will generate a Ly α photon, we apply an iterative algorithm. The expression

$$f_{\text{recycle},i} = \sum_f P_{if} f_{\text{recycle},f} \quad (3.12)$$

relates the conversion probability for the initial level i to the conversion probabilities of all

Table 3.1: Recycling fractions f_{recycle} and decay probabilities to the ground state, $P_{nP \rightarrow 1S}$.

possible lower levels f . The decay probabilities are calculated using equation (3.9). We then iterate from low to high n , calculating each f_{recycle} in turn.

In our particular case of an optically thick medium, we can ignore direct transitions to the ground state. These generate a Ly n photon, which will rapidly be reabsorbed and regenerate the nP state. Therefore, such decays will not affect the net population of photons or of excited states. We incorporate this into the calculation by setting $A_{nP \rightarrow 1S} = 0$ (Furlanetto et al., 2005).

Results for the lowest Lyman series transitions are summarised in Table 1 and plotted in Figure 3.3.⁴ These results are in agreement with those of Hirata (2006). At large n , the conversion fractions asymptote to $f_{\text{recycle}} \approx 0.36$ because nearly all cascades pass through lower levels. We emphasise again that the quantum selection rules forbid a Ly β photon from producing a Ly α photon.

n	f_{recycle}	$P_{nP \rightarrow 1S}$	n	f_{recycle}	$P_{nP \rightarrow 1S}$
			16	0.3550	0.7761
2	1	1	17	0.3556	0.7754
3	0	0.8817	18	0.3561	0.7748
4	0.2609	0.8390	19	0.3565	0.7743
5	0.3078	0.8178	20	0.3569	0.7738
6	0.3259	0.8053	21	0.3572	0.7734
7	0.3353	0.7972	22	0.3575	0.7731
8	0.3410	0.7917	23	0.3578	0.7728
9	0.3448	0.7877	24	0.3580	0.7725
10	0.3476	0.7847	25	0.3582	0.7722
11	0.3496	0.7824	26	0.3584	0.7720
12	0.3512	0.7806	27	0.3586	0.7718
13	0.3524	0.7791	28	0.3587	0.7716
14	0.3535	0.7780	29	0.3589	0.7715
15	0.3543	0.7770	30	0.3590	0.7713

⁴Code for calculating the conversion factors is available at <http://www.tapir.caltech.edu/~jp/cascade/>.

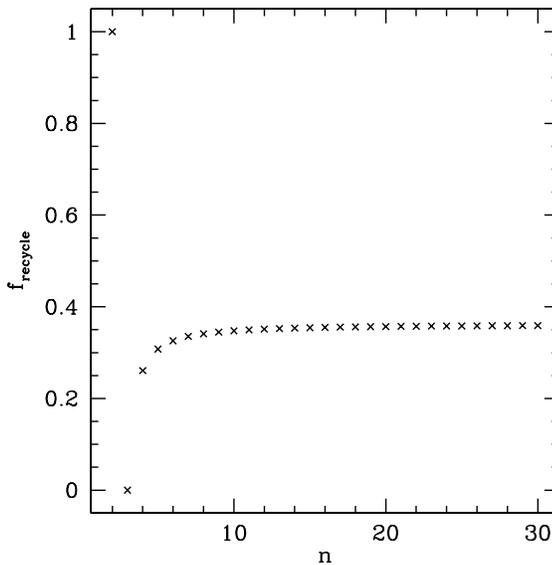


Figure 3.3: Recycling fractions for $\text{Ly}n$ photons. Note that the values level off at $f_{\text{recycle}} \approx 0.359$ and that none of the photons incident on the $\text{Ly}\beta$ resonance are converted into $\text{Ly}\alpha$ photons.

Finally we briefly comment on the two-photon decay from the $2S$ level (see Hirata, 2006, for a more detailed discussion). Selection rules forbid electric dipole transitions from the $2S$ level to the ground state, but the second order two-photon decay process can occur with $A_{\gamma\gamma} = 8.2 \text{ s}^{-1} \ll A_{2P \rightarrow 1S}$. At $z \lesssim 400$ the CMB flux density is sufficiently small that radiative excitations from the $2S$ level are negligible. Additionally, at the relevant densities collisional excitation to the $2P$ level is slow compared to the two-photon process (Breit & Teller, 1940). Consequently, the $2S$ level will preferentially decay via this two-photon process. These transitions may themselves affect T_s , because both the $2S$ and $1S$ levels have hyperfine structure and any imbalance in the decay constants here would affect the $1S$ populations. However, even without detailed calculations, we can see that the resultant coupling must be small. Cascades that do not generate $\text{Ly}\alpha$ must reach the $2S$ level, so the fraction of $\text{Ly}n$ photons that undergo two-photon decay is $f_{\gamma\gamma}(n) = 1 - f_{\text{recycle}}(n) \approx 0.64$ (See Table 1). Each such decay has $N_{\text{scat},\gamma\gamma} = 1$ because the resulting photons are not reabsorbed. This is much smaller than $N_{\text{scat},\alpha} \approx 10^6$. Consequently, only if the coupling per scattering were many orders of magnitude larger for two-photon decay than for $\text{Ly}\alpha$ scattering could this effect be significant.

3.5 The Ly α coupling around a source

We can see the effects of these recycling fractions on the Ly α coupling by considering the life of a photon emitted from a given source. The photon initially propagates freely, redshifting until it enters a Ly n resonance. Because the IGM is so optically thick, the photon will then scatter several times until a cascade converts it into a Ly α photon or two $2S \rightarrow 1S$ photons. In the latter case, the photons escape to infinity; in the former case, it scatters $\sim \tau$ times before redshifting out of the Ly α resonance. This establishes a series of closely-spaced horizons, because a photon entering the Ly n resonance at z must have been emitted below a redshift

$$1 + z_{\max}(n) = (1 + z) \frac{[1 - (n + 1)^{-2}]}{(1 - n^{-2})}. \quad (3.13)$$

The number of Ly n transitions contributing Ly α photons is thus a function of the distance from the source. These horizons imprint well-defined atomic physics onto the coupling strength by introducing a series of discontinuities into the Ly α flux profile of a source.

Thus the Ly α flux, J_α , arises from a sum over the Ly n levels, with the maximum n determined by the distance. The sum is ultimately truncated at $n_{\max} \approx 23$ to exclude levels for which the horizon lies within the HII region of a typical (isolated) galaxy, as only neutral hydrogen contributes to 21 cm absorption (BL05). The average Ly α background is thus

$$J_\alpha(z) = \sum_{n=2}^{n_{\max}} J_\alpha^{(n)}(z) = \sum_{n=2}^{n_{\max}} \int_z^{z_{\max}(n)} dz' f_{\text{recycle}}(n) \frac{(1+z)^2}{4\pi} \frac{c}{H(z')} \epsilon(\nu'_n, z'), \quad (3.14)$$

where ν'_n is the emission frequency at z' corresponding to absorption by the level n at z

$$\nu'_n = \nu_n \frac{(1+z')}{(1+z)}, \quad (3.15)$$

and $\epsilon(\nu, z)$ is the comoving photon emissivity (defined as the number of photons emitted per unit comoving volume, per proper time and frequency, at frequency ν and redshift z). To calculate $\epsilon(\nu, z)$, we follow the model of BL05

$$\epsilon(\nu, z) = \bar{n}_b^0 f_* \frac{d}{dt} F_{\text{gal}}(z) \epsilon_b(\nu), \quad (3.16)$$

where \bar{n}_b^0 is the cosmic mean baryon number density today, f_* is the efficiency with which gas is converted into stars in galactic halos (and with which Lyman-continuum photons escape

their hosts), and $F_{\text{gal}}(z)$ is the fraction of gas inside galaxies at z . We model the spectral distribution function of the sources $\epsilon_b(\nu)$ as a separate power law $\epsilon_b(\nu) \propto \nu^{\alpha_s-1}$ between Ly α and Ly β and between Ly β and the Lyman limit. In our calculations, we will assume Population III stars with spectral index $\alpha_s = 1.29$ between Ly α and Ly β , normalised to produce 4800 photons per baryon between Ly α and the Lyman limit, of which 2670 photons are emitted between Ly α and Ly β . In contrast, for Population II stars the numbers are 0.14, 9690, and 6520 respectively (BL05). In calculating F_{gal} , we use the Sheth & Tormen (1999) mass function dn/dm , which matches simulations better than the Press & Schechter (1974) mass function, at least at low redshifts. We assume that atomic hydrogen cooling to a virial temperature $T_{\text{vir}} \approx 10^4$ K sets the minimum halo mass. We will normalise f_* so that $x_\alpha = 1$ at $z = 20$; this yields $f_* = 0.16\%$ when we include the correct f_{recycle} .

To see what fraction of photons from a given source are converted into Ly α , we integrate $\epsilon_b(\nu)$ with the proper weighting by f_{recycle} . We find that $\bar{f}_{\text{recycle}} = 0.63, 0.72,$ and 0.69 for $\alpha_s(\text{Ly}\alpha - \text{Ly}\beta) = 1.29, 0.14,$ and -1.0 respectively (roughly corresponding to Pop. III stars, low metallicity Pop II stars, and quasars; see Zheng et al., 1997). The total flux is significantly less than if $f_{\text{recycle}} = 1$, as has been generally assumed before. Thus Ly α coupling will take place later if the proper atomic physics are included (typically $\Delta z \gtrsim 1$ for fixed source parameters; see also Hirata, 2006). Of course, this is only the average value, and around a given source there will be a distance dependence. A gas element that can only be reached by photons redshifted from below the Ly β resonance will see an effective $f_{\text{recycle}} = 1$. In contrast, a gas element very close to the source will have $f_{\text{recycle}} \approx 0.36$. This will be reflected in the brightness temperature power spectrum (see §3.6).

The Ly α flux profile of a galaxy with $M_{\text{gal}} = 3 \times 10^{10} M_\odot$ and our fiducial parameters at $z = 20$ is plotted in Figure 3.4. In our approximation, $J_\alpha \propto M_{\text{gal}}$. Thus obtaining $x_\alpha \geq 1$ at $r = 10$ Mpc requires a galaxy mass of $M_{\text{gal}} = 4.2 \times 10^{12} M_\odot$, corresponding to a 14σ fluctuation in the density field. Obviously, individual sources do not induce strong Lyman coupling on large scales. The conversion of photons from Ly γ to Ly α steepens the flux profile beyond the simple $1/r^2$ form. Notice that we have normalised f_* for each curve separately, so that $x_\alpha = 1$ at $z = 20$. Because setting $f_{\text{recycle}} = 1$ weights large n transitions more heavily (and hence small scales), that curve lies below the others at large r . The discontinuities occur at the Ly n horizons. In theory, their positions yield standard rulers determined by simple atomic physics. In practice, the weakness of the

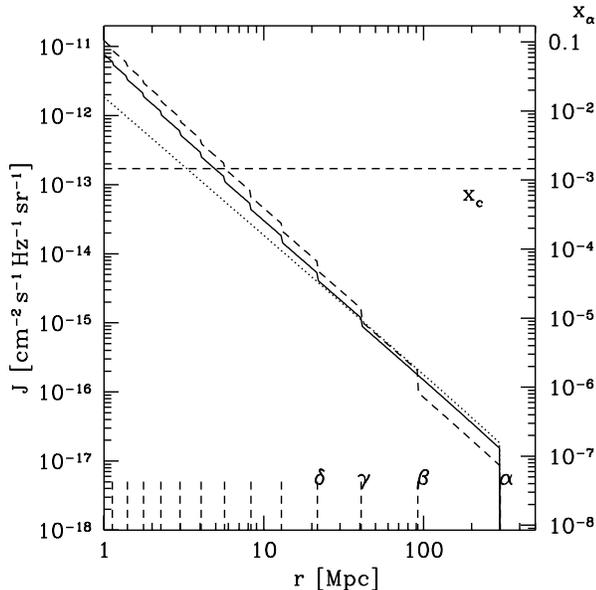


Figure 3.4: Flux profile of a galaxy of mass $M = 3 \times 10^{10} M_{\odot}$ at $z = 20$ as a function of comoving distance r . For comparison, we have plotted the flux profile assuming $f_{\text{recycle}} = 1$ (dashed curve), proper atomic physics (solid curve) and including only photons with $\nu_{\alpha} < \nu < \nu_{\beta}$ (dotted curve). Vertical lines along the lower axis indicate the horizons for the *Lyn* resonances. The horizontal dashed line shows the value of x_c at $z = 20$, illustrating the regime where collisional coupling dominates.

discontinuities, and the overlapping contributions of other nearby sources, makes it unlikely that these discontinuities will be observable for an isolated source (see also §3.6). Finally, we note that sharp discontinuities only occur if a photon undergoes a cascade immediately after entering a *Lyn* resonance and if the resulting $\text{Ly}\alpha$ photons redshift out of the $\text{Ly}\alpha$ resonance immediately. The former is certainly true, but the latter will affect the shape significantly (Loeb & Rybicki, 1999).

Additionally, photon horizons affect the radiation heating of the gas around a source. The cascades, which result from scattering of *Lyn* photons, deposit some of their radiative energy into the kinetic energy of the gas, and their total heating rate differs from “continuum” $\text{Ly}\alpha$ photons because they are injected as line photons (Chen & Miralda-Escudé, 2004). Thus the $\text{Ly}\alpha$ heating profile will differ from $1/r^2$. However, $\text{Ly}\alpha$ heating is typically much smaller than other sources, so this is unlikely to be important (Chen & Miralda-Escudé, 2004).

3.6 Brightness fluctuations from the first galaxies

In the previous section, we saw that the proper f_{recycle} affects the spatial distribution of x_α around each source. The most important manifestation of this will occur when the Wouthuysen-Field effect is just becoming important, around the time of the first galaxies (BL05). Those authors showed that Ly n transitions enhance the small-scale fluctuations in T_b , but they assumed that $f_{\text{recycle}} = 1$. We will show how the scale-dependent f_{recycle} modify this signal. It is possible to exploit the separation of powers to probe separately fluctuations that correlate with the density field and those, like Poisson fluctuations, that do not. We consider each in turn and compare to the results of BL05. We set $\delta_{x_{\text{HI}}} = 0$ throughout. We also assume that the IGM cools adiabatically, with no heat input from X-rays. Note that for ease of comparison with BL05, we do not incorporate the low-temperature corrections of Hirata (2006) (and in any case they are small in our example).

3.6.1 Density fluctuations

Density perturbations source x_α fluctuations via three effects (BL05). First, the number of galaxies traces, but is biased with respect to, the underlying density field. As a result an overdense region will contain a factor $[1 + b(z)\delta]$ more sources, where $b(z)$ is the (mass-averaged) bias, and will have a larger x_α . Next, photon trajectories near an overdense region are modified by gravitational lensing, increasing the effective area by a factor $(1 + 2\delta/3)$. Finally, peculiar velocities associated with gas flowing into overdense regions establish an anisotropic redshift distortion, which modifies the width of the region corresponding to a given observed frequency. These three effects may be represented using a linear transfer function $W(k)$ relating fluctuations in the coupling δ_{x_α} to the overdensity δ

$$\delta_{x_\alpha} \equiv W(k)\delta. \quad (3.17)$$

We compute $W(k)$ for a gas element by adding the coupling due to Ly α flux from each of the Ly n resonances (BL05)

$$W(k) = \frac{1}{x_\alpha} \sum_{n=2}^{n_{\text{max}}} \int_z^{z_{\text{max}}(n)} dz' \frac{dx_\alpha^{(n)}}{dz'} \frac{D(z')}{D(z)} \times \left\{ [1 + b(z')]j_0(kr) - \frac{2}{3}j_2(kr) \right\}, \quad (3.18)$$

where $D(z)$ is the linear growth function and the $j_l(x)$ are spherical Bessel functions of order l . The first term in brackets accounts for galaxy bias while the second describes velocity effects. The ratio $D(z')/D(z)$ accounts for the growth of perturbations between z' and z . The factor dx_α/dz converts from Ly α flux to the coupling. Each resonance contributes a differential coupling (see eq. 3.6)

$$\frac{dx_\alpha^{(n)}}{dz'} \propto \frac{dJ_\alpha^{(n)}}{dz'}, \quad (3.19)$$

with the differential comoving flux in Ly α from equation (3.14).

Because this correlates with the density field, it is easiest to observe via

$$P_{\mu^2}(k) = 2P_\delta(k) \left[\beta + \frac{x_\alpha}{\tilde{x}_{\text{tot}}} W(k) \right]. \quad (3.20)$$

The first term probes fluctuations in T_k and κ_{1-0} (all encoded in β). We show P_{μ^2} in Figure 3.5,⁵ contrasting cases that include only photons with $\nu_\alpha < \nu < \nu_\beta$, $\nu_\alpha < \nu < \nu_\delta$, and the entire Lyman continuum. For the latter two models, we show results with $f_{\text{recycle}} = 1$ and with the proper atomic physics. Note that each is separately normalised to $x_\alpha = 1$. The dotted curve isolates $2P_\delta\beta$, which clearly dominates on small scales. Note that we have applied two cutoffs to the power spectrum in this regime (BL05). The first is due to baryonic pressure, which prevents collapse on small scales. The second is the thermal width of the 21 cm line. Naoz & Barkana (2005) showed that this thermal cutoff displays a characteristic angular dependence, which, theoretically, allows fluctuations and the cutoff to be separated. To allow easy comparison with BL05, we do not include this refinement. Additionally, we expect power from the HII regions surrounding the sources to become important on scales smaller than the size of a typical HII region r_{HII} . We have marked this scale for an isolated galaxy in Figure 3.5, but note that Furlanetto et al. (2004) predict that the HII regions could be a factor of a few larger at these early times. On sufficiently large scales $kr \approx 0$, the second term in equation (3.20) dominates, and $W(k)$ is fixed by the source bias.

Figure 3.5 clearly shows that the Ly n resonances are important on intermediate scales. On large scales only the average flux matters, but as we move to smaller scales the higher- n levels become important. Figure 3.6 shows that the fractional reduction in $W(k)$ on small

⁵Our results for $\sqrt{P_{\mu^2}}$ are a factor of $\sqrt{2\pi^2} \approx 4.4$ greater than those of BL05, who made an error in their $P_\delta(k)$ normalisation (R. Barkana, private communication). Note that all of their density-induced fluctuation amplitudes should increase by a similar amount.

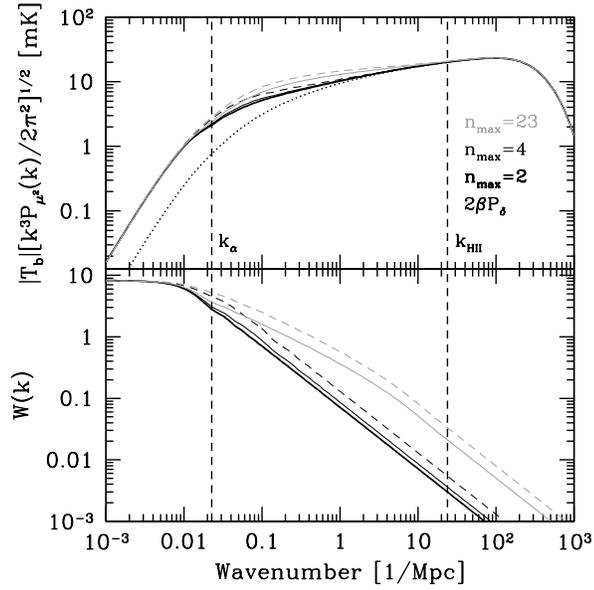


Figure 3.5: *Top panel:* P_{μ^2} power spectrum, which best illustrates the T_b fluctuations arising from density-sourced x_α . From bottom to top, the cases include only photons from Ly α to Ly β , from Ly α to Ly δ , and for $n \leq 23$. Dashed lines indicate $f_{\text{recycle}} = 1$ while solid lines use the proper conversion factors. The dotted line is $2\beta P_\delta$. Vertical lines indicate the scales corresponding to the Ly α horizon r_α and the HII region size r_{HII} . *Bottom panel:* Transfer function $W(k)$.

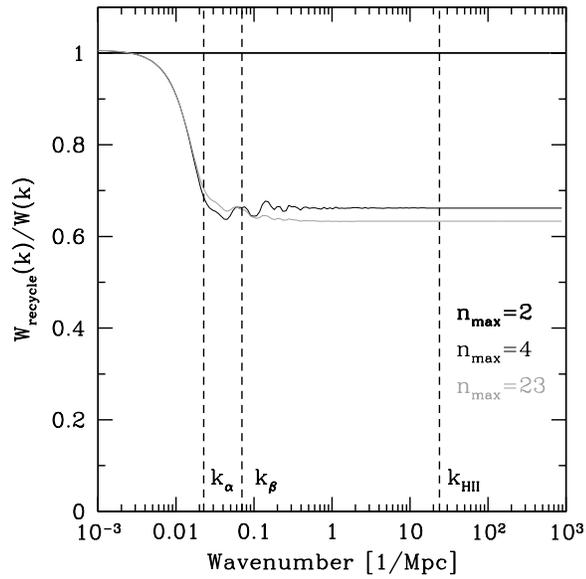


Figure 3.6: Ratio of $W(k)$ calculated using proper atomic physics to $W(k)$ with $f_{\text{recycle}} = 1$. From left to right, vertical lines indicate the scales associated with the Ly α and Ly β resonances and with r_{HII} . $W(k)$ displays small amplitude ripples, which arise from integrating an oscillating kernel (the spherical Bessel functions in equation 3.18) over finite extent (the Ly α horizons). Changing f_{recycle} modifies the phase of these ripples leading to the wiggles seen in $W_{\text{recycle}}(k)/W(k)$ on intermediate scales.

scales is ~ 0.63 . Although this is near \bar{f}_{recycle} for Pop. III stars, that is not the origin of this scaling. In equation (3.18), we can write

$$x_\alpha = \sum_n a_n f_{\text{recycle}}(n), \quad (3.21)$$

and

$$W(k) = \frac{1}{x_\alpha} \sum_n b_n f_{\text{recycle}}(n), \quad (3.22)$$

where the b_n and a_n are defined by reference to equation (3.18) and the integral of equation (3.19) respectively. The former care only about the local flux, but the latter are averaged over the entire Lyman continuum. If $f_{\text{recycle}} = \text{constant}$, they cancel out of $W(k)$ and are relevant only as an overall normalisation of x_α . In actuality, f_{recycle} is a function of n , reducing the power. We stress that this is because the $f_{\text{recycle}}(n)$ are essentially frequency dependent and so distort the flux profile about any isolated galaxy (see Fig. 3.4).

It might be hoped that the discontinuities in Figure 3.4 would leave a clear feature on the power spectrum, especially one associated with the loss of all photons entering the Ly β resonance. Such a feature, whose angular scale would be determined by simple atomic physics, could set a standard ruler that could be used to test variations in fundamental constants or to measure cosmological parameters. Sadly, as can be seen from Figure 3.5, there is no truly distinct feature. Still, the power does decline around k_α and measuring its shape can constrain the angular diameter distance: we find that the amplitude of P_{μ^2} changes by a few percent if the angular diameter distance changes by the same amount. However, such constraints would also require the astrophysical parameters to be known precisely, which will be difficult.

3.6.2 Poisson fluctuations

We turn now to brightness fluctuations uncorrelated with the underlying density perturbations, which can be extracted from the power spectrum because of the redshift space distortions (BL05). Specifically, if the number density of galaxies is small, then Poisson fluctuations can be significant. To calculate the correlation function from Poisson fluctuations, we again follow BL05 and consider the Ly α flux from sources within a volume element dV at two points A and B separated by a comoving distance l . The correlation function

takes the form

$$\xi_P(l) = \frac{2}{x_\alpha^2} \int_V dV \int_M \frac{dn(z'_A)}{dM} dM M^2 \frac{P(z'_A)}{r_A^2} \frac{P(z'_B)}{r_B^2} \frac{F_{\text{gal}}(z'_B)}{F_{\text{gal}}(z'_A)}, \quad (3.23)$$

where $z'_B = z'(r_B)$ is the redshift of a halo at a comoving distance r_B from a gas element at redshift z . In this expression, we integrate over a half volume such that $r_A < r_B$, with the factor of 2 accounting for the contribution of sources that are nearer to B. The factors $P(z')$ serve to normalise the flux from dV such that

$$dx_\alpha \equiv P(z') \frac{1}{r^2} \int_M M \frac{dn(z')}{dM} dM dV, \quad (3.24)$$

which makes explicit the expected $1/r^2$ dependence of the flux. Because of the finite speed of light, points A and B see the sources within dV at different stages in their evolution. Following BL05, we account for this with the last factor, which scales the source flux by the fraction of mass that has collapsed at the observed redshift. This ignores a possible dependence of the formation rate on halo mass, but in practice high redshift galaxies are highly biased and occur within a small mass range just above the minimum cooling mass, so this dependence will be weak. Equation (3.23) is easy to understand. For Poisson statistics, the variance of flux from a set of identical galaxies would be $\propto m_{\text{gal}}^2 n_{\text{gal}} V$; this must then simply be weighted by the flux reaching each of the two points. Note also that, contrary to the claims of BL05, $\xi_P \propto 1/f_d$, where f_d is the duty cycle of each galaxy, because the fluctuations are weighted by two powers of luminosity [$MP(z)$] but only one factor of the density.

The top panel of Figure 3.7 shows how the correlation function increases toward small scales. This is a result of the $1/r^2$ dependence of the flux, which weights the correlations to small scales. Including the Lyn resonances amplifies this, because the horizon scales skew the flux profile to small radii (see Figure 3.4). On large scales the correlation function decreases as the two points A and B share fewer sources. For $r > 2r_\alpha$ a single source cannot affect both points so $\xi_P = 0$. Including $f_{\text{recycle}}(n)$ reduces ξ_P , especially on the smallest scales, because it decreases the efficiency of coupling from level n . The bottom panel of Figure 3.7 shows that the suppression on small scales is $(0.63)^2 \approx 0.40$, because ξ_P depends on two powers of the flux. Note that, on large scales, ξ_P increases with the proper

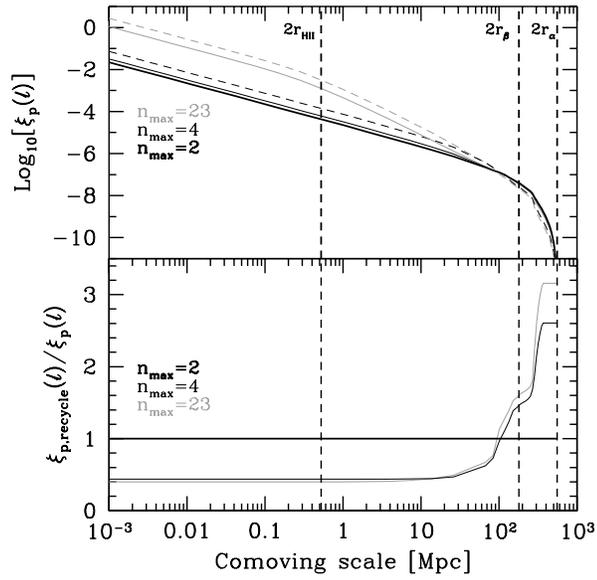


Figure 3.7: *Top panel*: Correlation function for the Poisson fluctuations. Line conventions are the same as in Figure 3.5. Vertical lines show $2r_\alpha$, $2r_\beta$, and $2r_{\text{HII}}$. *Bottom panel*: Ratio of the correlation functions assuming $f_{\text{recycle}}(n)$ and $f_{\text{recycle}} = 1$.

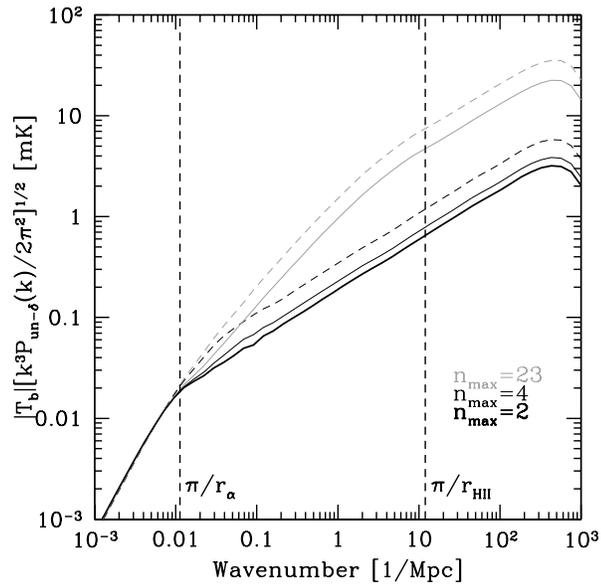


Figure 3.8: Power spectrum for the Poisson fluctuations. Line conventions are the same as in Figure 3.5. Vertical lines indicate the scales corresponding to $2r_\alpha$ and $2r_{\text{HII}}$.

f_{recycle} . This results from the way we have normalised to $x_\alpha = 1$, which reduces the flux of a given source on large scales as f_{recycle} increases (see Figure 3.4). As with $W(k)$, the scale dependence is weak except on large scales, where rapid changes occur at the appropriate horizons. Proper treatment of the recycling fractions is clearly necessary to understand the shape of ξ_P .

These features have similar effects on the power spectrum of fluctuations uncorrelated with the density fluctuations

$$P_{un-\delta}(k) \equiv P_{\mu^0} - \frac{P_{\mu^2}^2}{4P_{\mu^4}} = \left(\frac{x_\alpha}{\tilde{x}_{tot}} \right)^2 P_P(k), \quad (3.25)$$

as shown in Figure 3.8. On large scales, taking $f_{\text{recycle}} = 1$ slightly amplifies the power. On small scales, they significantly reduce the power by $\approx 60\%$. They also affect the shape of the power spectrum, especially near $k = \pi/r_\alpha$, where the lack of $\text{Ly}\beta \rightarrow \text{Ly}\alpha$ imprints a knee on the power spectrum. We also note that the sharp Lyn horizons imprint weak oscillations on the power spectrum, especially if $f_{\text{recycle}} = 1$ (though these will likely be smoothed by photon diffusion).

3.6.3 Nonlinearities in the Wouthuysen-Field coupling

To this point, we have used equation (3.7) to compute the brightness temperature fluctuations. This assumes that all the underlying perturbations are linear; obviously, at $x_\alpha = 1$, this may only be marginally satisfied. When the radiation background is large, the brightness temperature becomes insensitive to the coupling strength and P_{T_b} will be smaller than our estimate. When will such corrections become important? One obvious test is whether the typical fluctuation T_b is comparable to the maximum brightness temperature decrement between coupled and uncoupled gas, δT (i.e., if $T_s = T_k$ in eq. 1). But nonlinearities may be important even if this condition is not satisfied. A universe with discrete strongly coupled regions separated by uncoupled IGM could have small rms variations, even though nonlinearities are extremely important in fixing the brightness temperature of the strongly-coupled regions.

Instead we must look deeper at the nature of the fluctuations. First, note from Figure 3.4 that individual galaxies most likely provide only weak coupling: $x_\alpha \ll 1$ except near to the sources, at least if small galaxies (near the atomic cooling threshold) are responsible for

most of the radiation background. This is not surprising: because of Olber's paradox, each logarithmic radius interval contributes equally to the background flux in a homogeneous universe. The higher Lyman-series photons, together with the finite speed of light, also do not dramatically increase the weighting on nearby radii. Thus we expect a substantial fraction of the flux to come from large distances, where density fluctuations are weak. This immediately suggests that the density-dependent power spectrum described in Section 3.6.1 will not require substantial nonlinear corrections.

More quantitatively, the fluctuations become nonlinear when $\delta_{x_\alpha} = W(k)\delta(k) \gtrsim 1$; thus we require

$$W(k) \gtrsim 1/[\sigma(R)D(z)], \quad (3.26)$$

where $\sigma(R)$ is the typical density fluctuation on scale $R \sim 1/k$. Figure 3.5*b* shows, however, that $W(k)$ is of order unity only for $k \lesssim 0.1 \text{ Mpc}^{-1}$, where the density fluctuations are themselves tiny at these redshifts. Thus, we conclude that a linear treatment is adequate for computing the P_{μ^2} power spectrum, because it is primarily driven by large scale fluctuations.

The Poisson fluctuations in Section 3.6.2 are more problematic. By definition,

$$\xi_P(r_A - r_B) = \langle x_\alpha(r_A) x_\alpha(r_B) \rangle / x_\alpha^2 - 1. \quad (3.27)$$

Thus, on scales at which $\xi_P \gtrsim 1$, the radiation background near galaxy overdensities on this scale is considerably larger than its average value, indicating that nonlinear effects are important. In the particular model we have examined, ξ_P is large only on comoving scales $\lesssim 10 \text{ kpc}$, so nonlinear effects are again negligible. However, the amplitude of the Poisson fluctuations increases rapidly as the source density decreases: in models with fewer sources at $x_\alpha = 1$, or which strongly weight massive galaxies, nonlinearities may be important. A maximum value to the variance on any scale is $\delta T^2 Q(1 - Q) < 0.25 \delta T^2$, where Q is the volume filling factor of regions with $x_\alpha \gtrsim 1$. In particular, a linear treatment for sources with $f_\star \propto m^{2/3}$ can violate this limit on scales $k \gtrsim 1 \text{ Mpc}$ (see, e.g., Figs. 5 and 7 of BL05); in those cases the observed fluctuations can be much weaker than linear theory predicts (though it will also be non-gaussian).

3.7 Conclusions

In this paper, we have explored the effects on the spin-kinetic temperature coupling of photons that redshift into Lyman resonances. First, we considered the effect of direct coupling via resonant scattering of $\text{Ly}n$ photons. We showed that the possibility of cascades greatly reduces the number of times a $\text{Ly}n$ photon scatters before escaping. Consequently, the coupling is negligible and may be correctly ignored. A side effect of the reduced scattering rate is to make the $\text{Ly}n$ contribution to IGM heating extremely small, even if T_n is not in equilibrium with T_k .

Next we considered the increased $\text{Ly}\alpha$ flux that results from atomic cascades. Following the selection rules and transition rates, we calculated the probability that a $\text{Ly}n$ photon is converted into $\text{Ly}\alpha$ and showed that $f_{\text{recycle}} \rightarrow 0.36$ as n increases. This is significantly smaller than the value $f_{\text{recycle}} = 1$ usually assumed [e.g., by BL05]. For a typical Pop. III source spectrum, we showed that only 63% of the emitted photons will be converted into $\text{Ly}\alpha$, delaying the onset of coupling ($x_\alpha = 1$) for a given set of source parameters.

Incorporating the correct f_{recycle} modifies the flux profile of an individual source and reduces the coupling on small scales by about a factor of 3 for fixed source parameters. In addition, the cascade process imprints discontinuities onto the flux profile. Using the correct atomic physics reduces the amplitude of these discontinuities and removes one due to the $\text{Ly}\beta$ resonance. Unfortunately, their weakness is likely to frustrate attempts to use these discontinuities as a standard ruler.

We then recalculated the power spectra of BL05, incorporating the correct f_{recycle} . This showed a reduction in power of $\sim 37\%$ (Figure 3.5) on intermediate scales for density correlated fluctuations and of $\sim 64\%$ (Figure 3.8) on small scales for fluctuations uncorrelated with the density. It is possible to mimic this loss of power by changing the shape of the stellar spectrum. On small scales, a reduction in the star formation rate will produce a similar reduction in flux. Incorporating the proper $f_{\text{recycle}}(n)$ is thus crucial to correctly interpreting 21 cm observations near the time of first light. On the other hand, the effects that we have described become unimportant once $\text{Ly}\alpha$ coupling saturates so that $T_s \rightarrow T_k$. In this regime, fluctuations in the $\text{Ly}\alpha$ flux have little effect on the power spectrum and perturbations in the density and reionization fraction dominate.

Several experiments, including LOFAR, MWA, PAST, and the future SKA⁶, are aiming to detect 21cm fluctuations of the type we have discussed here. Hopefully, they will be able to study the first sources of light through their effect on the IGM around them. The details of Ly α coupling will determine the observability of this epoch.

Acknowledgments

We thank M. Kamionkowski for helpful discussions and also our anonymous referee for several helpful comments during the revision process. This work was supported in part by DoE DE-FG03-920-ER40701.

⁶See <http://www.skatelescope.org/>.

Chapter 4

The scattering of Lyman-series photons in the intergalactic medium

We re-examine scattering of photons near the Ly α resonance in the intergalactic medium (IGM). We first derive a general integral solution for the radiation field around resonance within the usual Fokker-Planck approximation. Our solution shows explicitly that recoil and spin diffusivity source an absorption feature, whose magnitude increases with the relative importance of recoil compared to Doppler broadening. This spectrum depends on the Ly α line profile, but approximating it with the absorption profile appropriate to the Lorentzian wings of natural broadening accurately reproduces the results for a full Voigt profile so long as the IGM temperature is less than ~ 1000 K. This approximation allows us to obtain simple analytic formulae for the total scattering rate of Ly α photons and the accompanying energy exchange rate. Our power series solutions converge rapidly for photons that redshift into the Ly α resonance as well as for photons injected at line center. We confirm previous calculations showing that heating through this mechanism is quite slow and probably negligible compared to other sources. We then show that energy exchange during the scattering of higher-order Lyman-series photons can be much more important than naively predicted by recoil arguments. However, the resulting heating is still completely negligible.

Originally published as Furlanetto and Pritchard, MNRAS, **372**, 1093 (2006).

4.1 Introduction

The radiative transfer of photons near the Ly α resonance is crucial to understanding the high-redshift intergalactic medium (IGM), both because it determines the spin temperature of the 21 cm transition (Wouthuysen, 1952; Field, 1958) and because it affects the thermal history (Madau et al., 1997; Chen & Miralda-Escudé, 2004).

The radiation field near this resonance has been examined a number of times in recent years. The earliest treatments ignored radiative transfer and assumed that the spectrum was featureless around the line. Chen & Miralda-Escudé (2004) were the first to solve (numerically) an approximate form of the radiative transfer equation in this context (following Basko 1981 and Rybicki & dell’Antonio 1994). They showed that, if photons redshift toward the resonance, the spectrum develops an asymmetric absorption feature. As we will see explicitly below, the absorption feature is sourced by recoil in the scattering process: each scattering deposits an average energy $\Delta E = (h\nu_\alpha)^2/(m_p c^2)$, where ν_α is the rest frequency of the Ly α line. Thus photons lose energy faster near the center of resonance, where they scatter more. To compensate for this increased “flow” speed, continuity requires that the amplitude of the background must decrease near resonance. This affects the scattering rate of Ly α photons and hence the spin temperature of the IGM. Hirata (2006) expanded on this method by showing how to account for the hyperfine structure of the Ly α line (see below).

An alternative to the numerical approach of Chen & Miralda-Escudé (2004) and Hirata (2006) is to approximate the spectrum analytically. This has a long history in resonant radiative transfer; Hummer & Rybicki (1992) summarize many of the advances. Of particular interest to our problem is the treatment of Grachev (1989), who derived an analytic solution for the spectrum around a resonant transition when recoil is included. The analytic solution was obtained by approximating the absorption profile using the form appropriate for scattering in the Lorentzian wings provided by natural broadening. This assumption is valid when the optical depth is extremely large and the Doppler broadening relatively small. Most recently, Chuzhoy & Shapiro (2006b) rediscovered this solution and applied it to the problem of Ly α transfer in the high-redshift IGM. In §4.2, we will show how these numeric and analytic solutions relate and study the validity of the analytic approximation. We also compute the radiation field in such a way that the role of recoil becomes obvious. We examine the resulting total scattering rate in §4.3 and show that the approximate form proposed by Chuzhoy & Shapiro (2006b) is a reasonably good match to the full numeric result. We also compute the colour temperature of the radiation field (relevant for the spin temperature of the 21 cm transition) in §4.4.

The line shape is also crucial for estimating the rate at which energy is transferred between the gas and the photon field. As described above, recoil during each scattering

deposits some energy in the gas. If this were the sole mechanism for energy exchange, the IGM would rapidly be heated above the cosmic microwave background (CMB) temperature (Madau et al., 1997). However, the absorption feature actually cancels almost all of this heating. Consider a photon on the blue side of the line. This will be preferentially scattered by an atom moving *away* from the photon (so that it appears closer to resonance). The atom will then re-emit the photon isotropically in its frame; in the IGM frame, the photon will therefore lose an energy $\sim h\Delta\nu_D$, where $\Delta\nu_D$ is the Doppler width of the transition. Photons that scatter on the red side, on the other hand, will tend to gain energy. The absorption feature develops so that this scattering “diffusivity” compensates for the recoil (i.e., so that more scattering occurs redward than blueward of the Ly α transition). The net energy transfer is therefore much slower than naively expected (Chen & Miralda-Escudé, 2004; Rybicki, 2006; Meiksin, 2006).

By employing their analytic approximation to the radiation field, Chuzhoy & Shapiro (2006a) took a step toward finding a simple solution for the net heating rate. In §4.5, we take their approach further by deriving a fully analytic solution for heating by photons redshifting into the Ly α resonance as well as an approximate solution for photons injected at line center (either through recombinations or cascades from higher Ly n transitions). This allows us to examine how the heating rate varies with IGM temperature and optical depth.

Of course, photons can redshift into any of the Ly n resonances in the IGM. After a few scatterings, these photons are destroyed through cascades to lower levels (Hirata, 2006; Pritchard & Furlanetto, 2006). The scattering rate is so small that recoil heating is negligible; however, all of the scatterings occur on the blue side of the line, so each deposits some fraction of the atom’s thermal energy in the gas as well. Chuzhoy & Shapiro (2006a) examined the analogous process in deuterium and found that it can provide relatively strong heating. In §4.6, we show that the heating rate for Ly n photons is tiny even when frequency drift is included, because the photons scatter so far in the blue wing of the line.

In our numerical calculations, we assume a cosmology with $\Omega_m = 0.26$, $\Omega_\Lambda = 0.74$, $\Omega_b = 0.044$, and $H = 100h \text{ km s}^{-1} \text{ Mpc}^{-1}$ (with $h = 0.74$), consistent with the most recent measurements (Spergel et al., 2006).

4.2 The radiation field near the Ly α resonance

We let J be the comoving angle-averaged specific intensity (in units of photons per area per steradian). The equation of radiative transfer is (neglecting atomic recoil for the moment)

$$\frac{1}{cn_H\chi_\alpha} \frac{\partial J}{\partial t} = -\phi(\nu) J + H\nu_\alpha \frac{\partial J}{\partial \nu} + \int d\nu' R(\nu, \nu') J(\nu') + C(t)\psi(\nu),$$

where n_H is the hydrogen density, $\sigma_\alpha(\nu) = \chi_\alpha\phi(\nu)$ is the absorption cross section, $\chi_\alpha = (\pi e^2/m_e c) f_\alpha$, f_α is the absorption oscillator strength, and $\phi(\nu)$ is the line profile. For our purposes, ϕ is given by the Voigt profile (which includes both collisional and natural broadening),

$$\phi(x) = \frac{a}{\pi^{3/2}} \int_{-\infty}^{\infty} dt \frac{e^{-t^2}}{a^2 + (x-t)^2}, \quad (4.1)$$

with $a = \Gamma/(4\pi\Delta\nu_D)$, Γ the inverse lifetime of the upper state, $\Delta\nu_D/\nu_0 = (2k_B T_K/mc^2)^{1/2}$ the Doppler parameter, ν_0 the line center frequency, T_K the gas temperature, and $x \equiv (\nu - \nu_0)/\Delta\nu_D$ the normalized frequency shift. The first term on the right-hand side of equation (4.1) describes absorption, the second the Hubble flow, and the third re-emission following absorption. The redistribution function $R(\nu, \nu')$ gives the probability that a photon absorbed at frequency ν' is re-emitted at frequency ν . The approximate form $R_{\text{II}}(\nu, \nu')$ (Henyey, 1941; Hummer, 1962), which assumes a Voigt profile with coherent scattering in the rest frame of the absorbing atom, is often used (see §4.6). We must, however, also include recoil (Basko, 1981) and, for exact calculations, spin exchange (Hirata, 2006; Chuzhoy & Shapiro, 2006b). The last term describes injection of new photons: C is the rate at which they are produced and $\psi(\nu)$ is their frequency distribution.

This integro-differential equation simplifies considerably if we assume that the background spectrum is smooth on the scale of the average frequency change per scattering (which is $\Delta x < 1$; see §4.6). In this Fokker-Planck approximation, equation (4.1) becomes (Rybicki & dell'Antonio, 1994)

$$\frac{d}{dx} \left\{ \phi(x) \frac{dJ}{dx} + 2[\eta' \phi(x) + \gamma'] J(x) \right\} + C\psi(x) = 0. \quad (4.2)$$

The coefficients γ' and η' depend on the scattering processes that are included in the redistribution function. Inserting the Hubble flow, recoil, and spin exchange, and further

assuming that $x \gg \nu_{21}/\Delta\nu_D$ (where ν_{21} is the frequency of the hyperfine transition),¹ these become (Hirata, 2006; Chuzhoy & Shapiro, 2006b)

$$\gamma' = \tau_{\text{GP}}^{-1}(1 + T_{\text{se}}/T_K)^{-1}, \quad (4.3)$$

$$\eta' = \eta \left(\frac{1 + T_{\text{se}}/T_S}{1 + T_{\text{se}}/T_K} \right) - (x + x_0)^{-1}, \quad (4.4)$$

where τ_{GP} is the total Gunn & Peterson (1965) optical depth of the Ly α transition, $\eta = (h\nu_0^2)/(mc^2\Delta\nu_D)$ is the mean (normalized) frequency drift per scattering from recoil (Basko, 1981), $x_0 \equiv \nu_0/\Delta\nu_D$ (this term enforces detailed balance; Rybicki 2006), T_S is the spin temperature of the 21 cm transition, and $T_{\text{se}} = (2/9)T_K\nu_{21}^2/\Delta\nu_D^2 = 0.40$ K. Before proceeding, we must note that equation (4.2) is not uniquely specified because there is some freedom in the drift and diffusivity imposed in the Fokker-Planck method. Other forms of the Fokker-Planck approximation have been examined by Meiksin (2006). Its utility for this problem has been verified numerically in some particular cases by Hirata (2006), but its accuracy for the general problem has not yet been fully explored (see also Meiksin 2006). We discuss its validity in more detail in §4.7.

The corrections for spin exchange, which are captured by the terms involving T_{se} , require some subtlety. Because Ly α transitions modify the ground-state hyperfine level populations (Wouthuysen, 1952; Field, 1958), the photons can also increase or decrease their frequency during each scattering by an amount corresponding to the energy defect of the 21 cm transition. This affects the flow rate of photons through the resonance (and the diffusivity) and hence the spectral shape. However, because the level populations themselves depend on the Ly α scattering rate, and because the mean energy exchange per scattering depends on the level populations, including spin exchange requires a simultaneous solution for T_S and the spectral shape (or, in practice, an iterative solution; Hirata 2006). For simplicity, we will neglect these corrections below. For clarity, we will therefore use $\gamma = \tau_{\text{GP}}^{-1}$ and η instead of their primed versions. To include spin exchange, one simply reverses this, as in Furlanetto et al. (2006) or Hirata (2006). The latter also shows the magnitude of these corrections (his Fig. 2): they are only a few percent except at $T_K \lesssim 1$ K (see also Tables 1 and 2 below).²

¹When $x \sim \nu_{21}/\Delta\nu_D$, the hyperfine splitting of the Ly α line cannot be ignored and a single line profile does not suffice; see Hirata (2006).

²Note that our definition of S_α below corresponds to \tilde{S}_α in Hirata (2006): see §4.4.

It is useful now to pause and note explicitly the scalings of the basic parameters of this problem; they will become useful later. We have $\Delta\nu_D \propto T_K^{1/2}$, so $a \propto T_K^{-1/2}$ and $\eta \propto T_K^{-1/2}$. The Sobolev parameter has $\gamma \propto (1+z)^{-3/2}$ in the high-redshift limit. Of course, spin exchange slightly modifies these scalings.

We will consider two sets of boundary conditions for equation (4.2). First, we let photons redshift into the resonance from large frequencies, with no injection term. To describe this we let $J_\infty > 0$ be the specific intensity as $x \rightarrow \infty$ and set $C = 0$. The second case allows injection at line center, so $C\psi(x) = C\delta(x)$,³ and sets $J_\infty = 0$. In this case, we define $J_{-\infty}$ to be the average intensity as $x \rightarrow -\infty$. In either scenario, equation (4.2) is easy to integrate once, leaving us with a first order ordinary differential equation. The formal solution is most transparently obtained by changing variables to (Hummer & Rybicki, 1992)

$$\sigma(x) = \int_0^x \frac{dx'}{\phi(x')}, \quad (4.5)$$

so that equation (4.2) becomes

$$\frac{dJ}{d\sigma} + 2(\eta\phi + \gamma)J = 2K, \quad (4.6)$$

where $K = \gamma J_\infty$ for the continuous case, $K = C$ for injected photons if $x < 0$, and $K = 0$ for injected photons with $x > 0$. Obviously

$$\exp \left[2\eta \int_0^\sigma \phi(\sigma') d\sigma' + 2\gamma\sigma \right] \quad (4.7)$$

is an integrating factor for this equation, from which the solution follows immediately. For injected photons with $x > 0$ (so that $K = 0$), it has the simple form

$$J(x) = J(0) \exp \left[-2\eta x - 2\gamma \int_0^x \frac{dx'}{\phi(x')} \right], \quad (4.8)$$

where $J(0)$ is determined by continuity.

A formal solution can also be written for $K > 0$, but in this case an alternate form is more physically illuminating. Here it is the absorption trough that is most interesting. To

³Even if the initial Lyman-series absorption occurs well blueward of line center, the Ly α photon that results from the cascade will be injected near line center because the atom passes through several intermediate states, each of which has a small natural width.

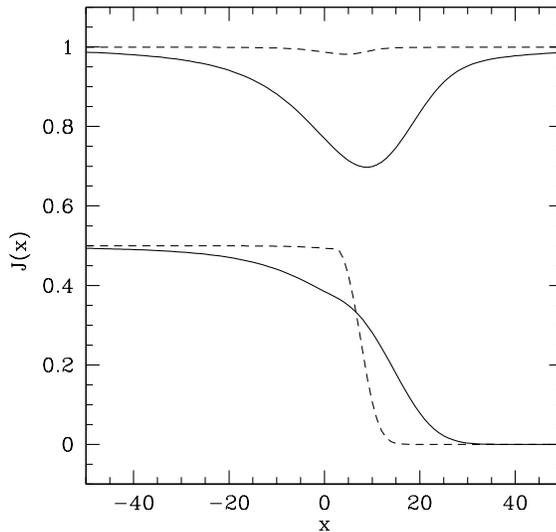


Figure 4.1: Background radiation field near the Ly α resonance at $z = 10$, assuming a Voigt line profile. The upper and lower sets are for photons redshifting from infinity and photons injected at line center, respectively. (The former are normalized to J_∞ ; the latter have $J_{-\infty} = 1/2$.) The solid and dashed curves take $T_K = 10$ and 1000 K, respectively.

isolate its properties, we define $\delta_J \equiv (J_\infty - J)/J_\infty$,⁴ note then that $\delta_J > 0$. The transfer equation takes the form

$$\phi \frac{d\delta_J}{dx} + 2(\eta\phi + \gamma)\delta_J = 2\eta\phi. \quad (4.9)$$

This has the same structure as the previous version, except that the sourcing term on the right-hand side depends on x . The same integrating factor yields the solution

$$\delta_J(x) = 2\eta \int_0^\infty dy \exp \left[-2\eta y - 2\gamma \int_{x-y}^x \frac{dx'}{\phi(x')} \right]. \quad (4.10)$$

This form makes it obvious that recoil sources the absorption spike. If the scattering were purely coherent, the gas and radiation field could not transfer any energy and the spectrum would remain flat (see, e.g., Hummer & Rybicki 1992). By sapping energy from each scattered photon, recoil increases the rate at which they redshift across the resonance. This increase in the “flow velocity” must be balanced by a corresponding decrease in the photon flux near the resonance.

⁴For injected photons, $J_\infty = 0$, of course; then we make the substitution $J_\infty \rightarrow J_{-\infty}$ in the definition. We will see that $J_\infty = J_{-\infty}$ for a redshifting continuum.

We show some example spectra in Figure 4.1, assuming that $\phi(x)$ has a Voigt profile (see also Chen & Miralda-Escudé 2004). The upper curves assume that photons redshift into resonance from infinity; as expected, an absorption feature develops. It deepens at small temperatures, because, in that case, the energy lost from recoil is large compared to the energy lost in each scattering (or η is relatively large). The lower curves assume injection at line center. In this case, the spectrum spreads to large positive x when T_K decreases.

This numerical solution is, of course, identical to those presented by Chen & Miralda-Escudé (2004) and Hirata (2006), once the appropriate line profiles, drifts, and diffusivities are inserted. It is also a more general form of the solutions provided by Hummer & Rybicki (1992) (who neglected the recoil term) and Chuzhoy & Shapiro (2006b). The latter implicitly made the approximation (following Chugai 1980, 1987; Grachev 1989) that $\phi(x) \approx a/(\pi x^2)$, which is only accurate at $|x| \gg 1$. We will refer to this as the “wing” approximation for convenience. This approximation allows the integrals over ϕ^{-1} to be performed analytically (Grachev, 1989; Chuzhoy & Shapiro, 2006b). For injected photons with $x > 0$, the solution is

$$J(x) = J(0) \exp\left(-2\eta x - \frac{2\pi}{3} \frac{\gamma x^3}{a}\right), \quad (4.11)$$

while for a flat background or injected photons with $x < 0$,

$$\delta_J(x) = 2\eta \int_0^\infty dy \exp\left[-\frac{2\pi\gamma}{3a}(y^3 - 3y^2x + 3yx^2) - 2\eta y\right]. \quad (4.12)$$

This explains the discrepancy between the existing numeric and analytic results: the latter do not apply near the Doppler core of the profile.

Figure 4.2 shows the ratio of the approximate analytic solutions of Grachev (1989) and Chuzhoy & Shapiro (2006b) to the exact spectra (computed with a Voigt profile).⁵ When the temperature is small (solid and dotted curves), the approximation is an excellent one. However, it begins to break down at large temperatures: for example, in the continuous case with $T_K = 1000$ K, it underpredicts δ_J by $\sim 10\%$ at the center of the absorption spike. This is because the effective natural width decreases with temperature, so the thermal broadening becomes more important in higher-temperature gas. Note as well that the deviation has a non-trivial shape. This is because the character of the wing approximation

⁵Or, more precisely, the exact solution within the Fokker-Planck approximation.

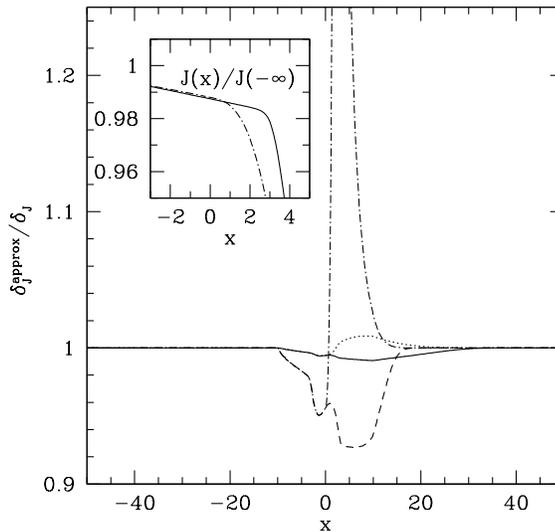


Figure 4.2: Ratio of δ_J in the “wing” approximation to the exact results (using a Voigt profile). The solid and dashed curves assume a flat background spectrum and take $T_K = 10$ and 1000 K, respectively. The dotted and dot-dashed curves assume injection at line center, with $T_K = 10$ and 1000 K, respectively. In this case, we set $\delta_J = J/J_{-\infty}$ for $x > 0$. The inset shows a closeup of J near resonance for injected photons at $T_K = 1000$ K; the solid and dot-dashed curves show the exact and approximate solutions, respectively. All curves assume $z = 10$.

changes depending on whether x is less than or greater than unity (that is, $\phi \rightarrow \infty$ when $x \rightarrow 0$ in the wing approximation). The deviation is worst for injected photons, especially at high temperatures. The physical effects are still small, however: in the inset we show a closeup of the spectrum itself near $x = 0$. We see that, in the wing approximation, the decline at $x > 0$ occurs a bit earlier. The fractional deviation is therefore large, but only over a limited range of frequencies (and only outside the line core).

Overall, we find that the wing approximation is an excellent one. Below we will use this analytic form to study the scattering and heating rates, extending the approach of Chuzhoy & Shapiro (2006b,a).

4.3 The Ly α scattering rate

The total rate at which Ly α photons scatter (per hydrogen atom) is

$$P_\alpha = 4\pi\chi_\alpha \int_{-\infty}^{\infty} d\nu J(\nu)\phi(\nu), \quad (4.13)$$

where J is now in proper units. Because each scattering can exchange hyperfine states, this rate is crucial for determining the spin temperature of the 21 cm transition in the IGM (Wouthuysen, 1952; Field, 1958; Madau et al., 1997; Chen & Miralda-Escudé, 2004; Hirata, 2006; Chuzhoy & Shapiro, 2006b). The Wouthuysen-Field coupling strength can be written as (e.g., Furlanetto et al. 2006)⁶

$$x_\alpha = \frac{16\pi\chi_\alpha J_\infty}{27A_{10}} \frac{T_\star}{T_\gamma} S_\alpha, \quad (4.14)$$

where $A_{10} = 2.85 \times 10^{-15} \text{ s}^{-1}$ is the spontaneous emission coefficient of the 21 cm transition, $T_\star = 0.068 \text{ K}$ is the energy defect of that transition, T_γ is the CMB temperature, and

$$S_\alpha \equiv \int_{-\infty}^{\infty} dx \phi(x) \frac{J}{J_\infty} \quad (4.15)$$

depends only on the *shape* of the background spectrum. Note that $S_\alpha < 1$, because recoil always induces an absorption feature.

In general, S_α must be computed numerically; even in the wing approximation, there is no closed-form analytic solution. However, recall that $\phi(x)$ is sharply peaked around $x = 0$, while J varies slowly near resonance (even in the injected case). Thus we can approximate $J \approx J(0)$ everywhere inside the integral; from the normalization of ϕ we thus have

$$1 - S_\alpha \approx \delta_J(0). \quad (4.16)$$

In the wing approximation, this is easily computed from equation (4.12):

$$1 - S_\alpha \approx \frac{4\alpha}{9} \left[3^{2/3} \pi \text{Bi} \left(-\frac{2\alpha}{3^{1/3}} \right) + (3\alpha^2) {}_1F_2 \left(1; \frac{4}{3}, \frac{5}{3}; -\frac{8\alpha^3}{27} \right) \right], \quad (4.17)$$

$$\approx \frac{4\pi}{3\sqrt{3}\Gamma(2/3)} \alpha - \frac{8\pi}{3\sqrt{3}\Gamma(1/3)} \alpha^2 + \frac{4}{3} \alpha^3 + \dots, \quad (4.18)$$

⁶For injected photons, one must substitute $J_\infty \rightarrow J_{-\infty}$. Note as well that J_∞ must be in proper units.

where $\text{Bi}(x)$ is an Airy function, ${}_1F_2$ is a hypergeometric function, and

$$\alpha = \eta \left(\frac{3a}{2\pi\gamma} \right)^{1/3} = 0.717 T_K^{-2/3} \left(\frac{10^{-6}}{\gamma} \right)^{1/3}, \quad (4.19)$$

where T_K is in degrees Kelvin; note that the second equality is not exact when spin exchange is important (which requires the replacements $\gamma \rightarrow \gamma'$ and $\eta \rightarrow \eta'$) or when the correction for detailed balance is significant. When α is small, we therefore have $(1 - S_\alpha) \propto T_K^{-2/3} \tau_{\text{GP}}^{1/3}$. This scaling gives some intuition for how the coupling strength varies in the IGM. As in Figure 4.1, the absorption spike becomes less and less significant as T_K increases; thus we must have $S_\alpha \rightarrow 1$ (its value without recoil) in a warm IGM. The perturbation increases with optical depth because that increases the number of scatterings (and hence the energy loss due to recoil).

We show the dependence of $(1 - S_\alpha)$ on temperature in Figure 4.3 and the dependence on the Sobolev parameter (or optical depth) in Figure 4.4. The thick curves show the numeric solution for a Voigt line profile and for a continuous background spectrum, which we denote S_c . The case with photons injected at the line center has a nearly identical scattering integral, because $\delta_J(0)$ is the same in the two cases; only at high temperatures does the structure around resonance matter. The thin curves show the first-order (in α) approximation of equation (4.18). We see that this provides an excellent match at $T_K \gtrsim 10$ K, especially when γ is relatively large (i.e., at lower redshifts).

Note that we have actually made three approximations here: (i) a constant J across the line, (ii) the wing approximation, and (iii) the small α approximation. The culprit at small T_K is the third. Here α is large and the power series approximation breaks down. However, even including just terms up to α^3 dramatically improves the estimate, with errors $\lesssim 10\%$ so long as $T_K > 2$ K. This demonstrates that the first approximation is an excellent one here: at such small temperatures, $\phi(x)$ is extremely sharply peaked. The second is equally good. Chuzhoy & Shapiro (2006b) proposed the fit

$$\delta_J(0) \approx 1 - \exp(-1.79\alpha), \quad (4.20)$$

which retains the first-order behavior of $\delta_J(0)$ at small α (and hence is reasonably accurate) and fits the behavior for $\alpha \sim 1$ much better. The thin dot-dashed curve in Figure 4.4 shows

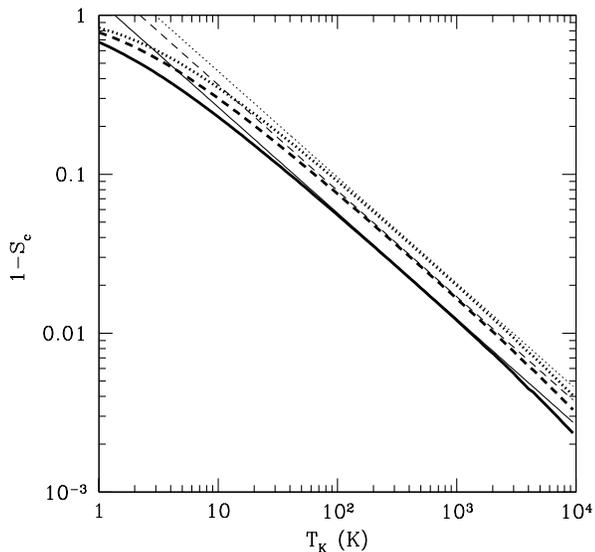


Figure 4.3: Scattering integral as a function of IGM temperature. The thick solid, dashed, and dotted curves show $(1 - S_c)$ for a Voigt profile at $z = 10, 20,$ and 30 . The thin curves show the corresponding quantities using only the first-order term in equation (4.18).

how well this approximation does at $T_K = 1$ K; it typically differs from the exact solution by $\sim 5\%$.

The overall agreement worsens at large temperatures as well. Here α is small, so the power series in equation (4.18) converges rapidly and approximation (iii) is excellent. The problem lies instead with the other two. As we have seen, the wing approximation breaks down once T_K exceeds ~ 1000 K. This causes up to a 10% underestimate of $\delta_J(0)$. At the same time, approximation (i) breaks down and the region around resonance starts to contribute to the scattering integral. This causes a $\lesssim 20\%$ overestimate of $(1 - S_c)$ compared to the exact result; fortunately, these two effects partially cancel.

In summary, the fit proposed by Chuzhoy & Shapiro (2006b) (in eq. 4.20) is an excellent approximation (within the Fokker-Planck formalism) to S_α unless high accuracy is required. However, we emphasize that, in order to include spin exchange properly, one must still use an iterative procedure (Hirata, 2006).

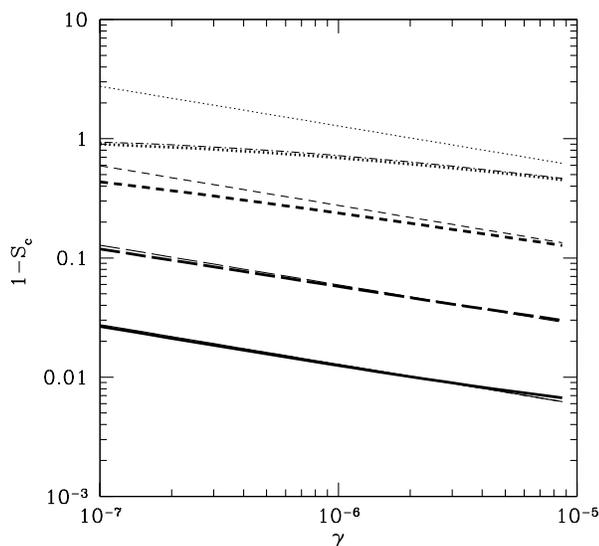


Figure 4.4: Scattering integral as a function of $\gamma = \tau_{\text{GP}}^{-1}$. The thick curves show $(1 - S_c)$ computed numerically for a Voigt profile, while the thin curves show the corresponding quantities using only the first-order term in equation (4.18). The dotted, short-dashed, long-dashed, and solid curves take $T_K = 1, 10, 10^2,$ and 10^3 K, respectively. The thin dot-dashed curve shows the approximate form proposed by Chuzhoy & Shapiro (2006b) for $T_K = 1$ K.

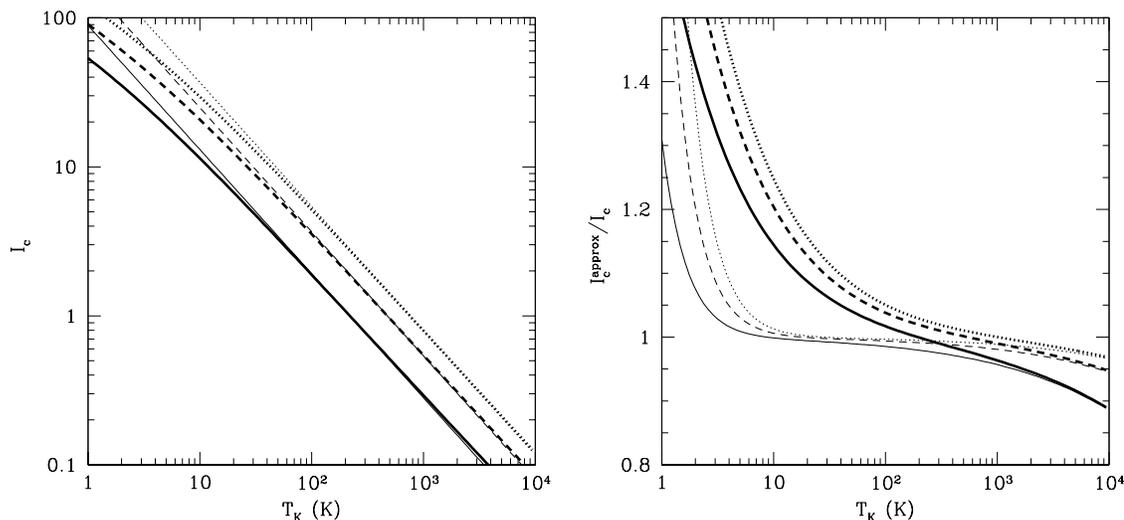


Figure 4.5: Heating integral for continuous injection. *Left panel:* The thick solid, dashed, and dotted curves show I_c for a Voigt profile at $z = 10, 20,$ and 30 . The thin curves show the corresponding quantities using only the first-order term in equation (4.30). *Right panel:* Ratio of the power series approximation to I_c (using the wing approximation) to the exact value. The thick and thin curves retain terms to order β and β^3 , respectively.

4.4 The colour temperature

The Ly α radiation field couples the spin temperature T_S to an effective colour temperature T_c , defined as (Rybicki, 2006)

$$\frac{h}{k_B T_c} \equiv -\frac{d \ln n_\nu}{d\nu}, \quad (4.21)$$

where $n_\nu = c^2 J / 2\nu^2$ is the photon occupation number. The spin temperature is then determined by (Field, 1958; Madau et al., 1997)

$$T_S^{-1} = \frac{T_\gamma^{-1} + x_c T_K^{-1} + x_\alpha T_c^{-1}}{1 + x_c + x_\alpha}, \quad (4.22)$$

where T_γ is the CMB temperature and x_c is the collisional coupling coefficient (see Furlanetto et al. 2006 and references therein). Of course, because of the non-trivial spectral shape near the Ly α resonance, T_c is actually a function of frequency; it should be harmonically averaged across the line profile to compute the effective coupling temperature (Chen & Miralda-Escudé, 2004; Meiksin, 2006). However, that makes only a small difference because the spectrum is so smooth (for the same reasons that eq. 4.16 is a good approximation).

At resonance, we can solve equation (4.2), with spin exchange included, to obtain (for photons that redshift into resonance)

$$\frac{h}{k_B T_c} = \frac{2\eta'}{\Delta\nu_D} - \frac{2\gamma'}{\phi_0 \Delta\nu_D} \frac{J_\infty - J}{J}, \quad (4.23)$$

where the 0 subscript indicates evaluation at $x = 0$. The second term describes the deviation sourced by the Hubble flow. It is generally small (and formally vanishes in the wing approximation) but is easily included; the result is (Chuzhoy & Shapiro, 2006b)

$$T_c = T_K \left[\left(\frac{1 + T_{se}/T_S}{1 + T_{se}/T_K} \right) - \frac{\gamma'}{\eta\phi_0} \frac{\delta_{J,0}}{1 + \delta_{J,0}} \right]^{-1}. \quad (4.24)$$

Ignoring the Hubble flow term, this matches the fit to numeric results proposed by Hirata (2006), provided that the correction from spin exchange is small. The Hubble flow term vanishes at small temperatures because then $\eta \gg \gamma'$; it also vanishes at high temperatures because then $\delta_{J,0}$ is small. It is never greater than $\sim 10^{-4}$. The spin exchange correction can be much larger.

Note that, because T_c itself depends on T_S , equation (4.22) is an implicit equation for

the spin temperature and must usually be solved simultaneously with the spectral shape and scattering rate. However, when $T_K, T_S \gg 1$ K, the corrections are small (Hirata, 2006).

4.5 Heat exchange from Ly α scattering

4.5.1 Continuous background

The rate at which the radiation field deposits energy in the gas (per unit volume) is (Chen & Miralda-Escudé, 2004)

$$\epsilon_\alpha = \frac{4\pi H h \nu_0}{c} \int_{-\infty}^{\infty} d\nu (J_\infty - J), \quad (4.25)$$

where we have assumed $\nu \approx \nu_0$ across the absorption feature and J is again in proper units. The physical interpretation of this form is straightforward: in the absence of scattering, the absorption feature would redshift away to infinity. To keep it in place, the photons must lose energy at the rate given in equation (4.25). More formally, it can be derived from the average energy exchange per scattering (Chuzhoy & Shapiro, 2006a) through integration by parts and the use of equation (4.2).

Thus the heating rate depends on

$$I_c = \int_{-\infty}^{\infty} dx \delta_J(x), \quad (4.26)$$

$$= 2\eta \int_0^{\infty} dy e^{-2\eta y} \int_{-\infty}^{\infty} dx \exp \left[-2\gamma \int_{x-y}^x \frac{dx'}{\phi(x')} \right]. \quad (4.27)$$

This cannot be done in closed form for an arbitrary line profile, but the accuracy of the wing approximation makes it extremely useful in understanding the solution.⁷ In this case, both integrals can be done analytically, yielding

$$I_c = \left(\frac{4}{\pi} \right)^{-1/6} \pi^{3/2} \left(\frac{a}{\gamma} \right)^{1/3} \beta [\text{Ai}^2(-\beta) + \text{Bi}^2(-\beta)], \quad (4.28)$$

where

$$\beta = \eta \left(\frac{4a}{\pi\gamma} \right)^{1/3} = 0.99 T_K^{-2/3} \left(\frac{\gamma^{-1}}{10^6} \right)^{1/3}, \quad (4.29)$$

⁷Note that Chuzhoy & Shapiro (2006a) calculated the heating rate numerically both in the wing approximation and using the full Voigt profile.

and $\text{Ai}(x)$ and $\text{Bi}(x)$ are the Airy functions and T_K is in degrees Kelvin. Note that this solution is *exact* within the wing approximation (although including spin exchange requires $\gamma \rightarrow \gamma'$ and $\eta \rightarrow \eta'$, so the second equality in eq. 4.29 is only approximate; it also ignores the detailed balance correction).

We can again find a simple and useful approximation by expanding in powers of β . We find

$$I_c \approx 3^{1/3} \left(\frac{2\pi}{3}\right)^{5/3} \left(\frac{a}{\gamma}\right)^{1/3} \left[\frac{\beta}{\Gamma^2(2/3)} - \frac{3^{1/3}\beta^2}{\Gamma(1/3)\Gamma(2/3)} + \frac{3^{2/3}\beta^3}{\Gamma^2(1/3)} + \dots \right]. \quad (4.30)$$

Thus, we see $I_c \propto T_K^{-5/6} \gamma^{-2/3}$; because $\gamma = \tau_{\text{GP}}^{-1} \propto (1+z)^{-3/2}$ at high redshifts, we expect $I_c \propto (1+z)$ at fixed temperature. The heat input per atom per Hubble time (at constant J_∞) is therefore $\Delta T \propto H\Delta\nu_D I_c / (n_{\text{HI}}H) \propto T_K^{-1/3} (1+z)^{-2}$. These scalings are close to those estimated by Chuzhoy & Shapiro (2006a) from their numerical results.

We show our solution for I_c as a function of T_K in the left panel of Figure 4.5 and as a function of γ in Figure 4.6. In each of these panels, the thick curves use the full Voigt profile, while the thin curves use the first-order term (in β) of equation (4.30); we expect the latter to be valid when $T_K \gtrsim 100$ K. The right panel of Figure 4.5 shows the ratio of the approximate and exact solutions; here the thick curves retain only the lowest-order term, while the thin curves include terms up to β^3 : these are necessary for $T_K \lesssim 10$ K. As before, the expansion in equation (4.30) converges rapidly at higher temperatures, but the wing approximation begins to break down.

Obviously the predicted scalings are reasonably accurate; the heating rate decreases with temperature (because recoil becomes relatively inefficient) and increases with τ_{GP} (along with the scattering rate). The higher-order terms, and the Voigt profile, slightly decrease the dependence on these parameters. As shown by Chen & Miralda-Escudé (2004), $\text{Ly}\alpha$ heating is probably slow compared to other processes, and the wing approximation (in the full analytic expression for small temperatures and the power series form otherwise) should be adequate for most purposes.

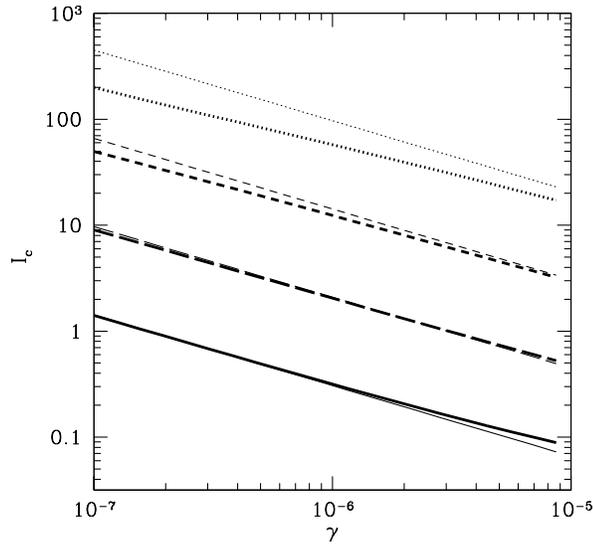


Figure 4.6: Heating integral for continuous injection. The thick curves show I_c computed numerically for a Voigt profile, while the thin curves show the corresponding quantities using only the first-order term in equation (4.30). The dotted, short-dashed, long-dashed, and solid curves take $T_K = 1, 10, 10^2, \text{ and } 10^3$ K, respectively.

4.5.2 Injection at line center

For photons injected at the line center, a similar exercise shows that the relevant integral is (Chen & Miralda-Escudé, 2004)

$$I_i = \int_{-\infty}^0 dx \delta_J(x) - \int_0^{\infty} dx \frac{J(x)}{J_{-\infty}}. \quad (4.31)$$

Again, we work in the wing approximation to gain some intuition. The second integral can be written in closed form; the first is

$$\int_{-\infty}^0 dx \delta_J = \frac{\eta}{\sqrt{2}} \sqrt{\frac{a}{\gamma}} \int_0^{\infty} \frac{dy}{\sqrt{y}} \exp\left(-\frac{\pi\gamma}{6a} y^3 - 2\eta y\right) \operatorname{erfc}\left(\sqrt{\frac{\pi\gamma}{2a}} y^3\right). \quad (4.32)$$

Unfortunately, the complementary error function prevents a closed form solution. However, note that the exponential term implies that the integral is dominated by the region where the argument of the error function is small. Expanding it to lowest-order, we then obtain a power series solution in β :

$$I_i \approx \left(\frac{a}{\gamma}\right)^{1/3} \sum_{i=0}^{\infty} A_i \beta^i. \quad (4.33)$$

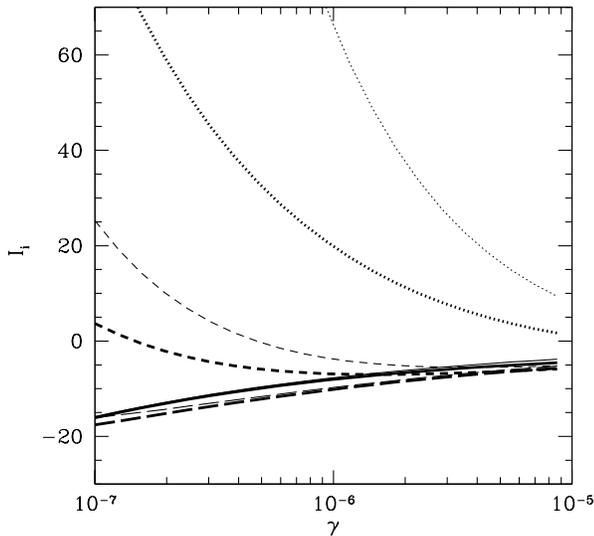


Figure 4.7: As Fig. 4.6, but for injection at line center. Here the approximate versions include terms up to order β .

The first few terms have $(A_0, A_1, A_2) = (-0.6979, 2.5424, -2.5645)$. Retaining only the zeroth-order term, the scaling with γ and T_K is again close to that proposed by Chuzhoy & Shapiro (2006a) at $T_K \gtrsim 100$ K.

Figures 4.7 and 4.8 show I_i for the same parameters as in Figures 4.5 and 4.6; again we compare the approximate form with the exact solution (including the full Voigt profile). In this case the dependence on both T_K and γ is considerably more complicated. Most interestingly, injected photons can both heat the gas (when $T_K \lesssim 10$ K) and cool it. Physically, cooling can occur because more photons scatter on the red than the blue side of the line; in such events, the re-emitted photon generally has a higher energy in the IGM frame and so removes heat from the gas. In the high-temperature regime ($T_K \gtrsim 100$ K), the cooling rate falls slightly when γ decreases and when T_K increases. At small temperatures, the exchange switches to heating because the feature is so broad compared to the $\Delta x \sim 1$ frequency change per scattering.

We also show the approximate form (eq. 4.33) in these panels (note that we must include β^0 and β^1 terms). It is substantially less accurate at first-order in β , only approaching the exact solution at $T_K \gtrsim 200$ K; shortly thereafter, the Voigt profile becomes significant. However, the thin curves in the right panel of Figure 4.8 show that carrying the series

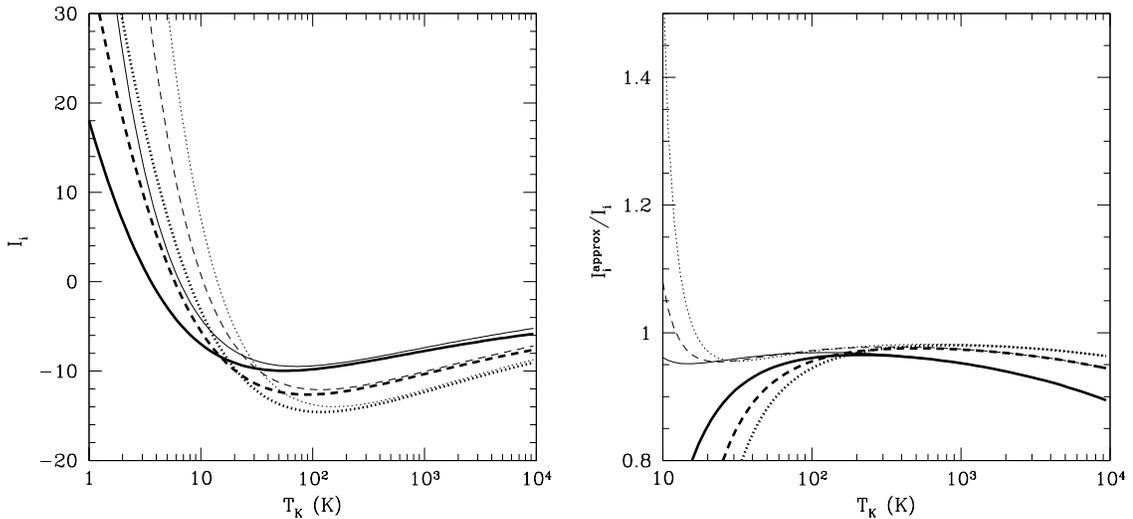


Figure 4.8: As Fig. 4.5, but for injection at line center. In the right panel, the approximate versions include terms up to order β (thick curves) and β^2 (thin curves).

expansion to β^2 is quite accurate throughout the range $T_K \gtrsim 10$ K. In the injected case, the wing approximation is less useful because no analytic solution exists. Thus we recommend numerical integration of equation (4.31) when high accuracy is required (especially at small temperatures).

4.6 Scattering of Ly α photons

Consider a photon that redshifts into a Ly α line (with frequency ν_n) at redshift z_r ; its frequency at redshift z is therefore $\nu_z = \nu_n[(1+z)/(1+z_r)]$. The accumulated optical depth it has traversed by that point is

$$\tau(z) = \int_z^\infty dz \frac{\sigma(\nu_z)n_{\text{HI}}(z)c}{(1+z)H(z)}, \quad (4.34)$$

where σ refers to the line of interest. The Gunn & Peterson (1965) optical depth is of course the *total* optical depth experienced by such a photon, or $\tau(z \ll z_r)$.

We are interested in determining the surface at which such a photon will *first* scatter. For the extremely optically thick Lyman-series lines of the IGM, this first scattering occurs far in the wings of the line, so we can set $\phi(x) \approx a/(\pi x^2)$. Further assuming the high-redshift

limit [$H(z) \propto (1+z)^{3/2}$] and letting $\Delta z \equiv z - z_r \ll z_r$, we have

$$\tau(z) \approx \frac{n_0^c \chi_n}{H_0 \sqrt{\Omega_0}} \frac{a}{\pi} \frac{(1+z_r)^{5/2}}{\Delta z \nu_n^2}, \quad (4.35)$$

where χ_n is evaluated for the line of interest. Re-expressing Δz in terms of x , we find

$$x(\tau) \approx \frac{1650}{\tau T_K^{1/2}} \left(\frac{\nu_\beta}{\nu_n} \right)^4 \left(\frac{A_n}{A_\beta} \right) \left(\frac{\Gamma_n}{\Gamma_\beta} \right) \left(\frac{1+z_r}{20} \right)^{3/2}, \quad (4.36)$$

where we have normalized ν_n , the spontaneous emission coefficients A_n , and the inverse lifetimes Γ_n to the values appropriate for Ly β . Higher Lyman-series photons have significantly longer lifetimes and hence scatter nearer line center; for example, Ly ϵ photons have a coefficient ≈ 18 .

By setting $\tau = 1$ in equation (4.36), we see immediately that the first scattering occurs well blueward of resonance; we will denote this location x_1 . Ly α photons cannot be destroyed during scattering (except, of course, in the exceedingly unlikely event that a collision occurs while the atom is excited), so this first scattering limit has little physical interest. However, higher Lyman-series photons can be destroyed, because the excited state can cascade to an intermediate level. The destruction probabilities per scattering are compiled by Hirata (2006) and Pritchard & Furlanetto (2006); they are $\sim 10\%$ for Ly β and $\sim 20\%$ for higher-level transitions. Thus, each such photon scatters only a few times before vanishing; so long as they remain in the wings, the k th scattering will occur at $x_k \approx x_1/k$.

Because the photons are far out on the blue wing during each of these scattering events, they will deposit some fraction of their energy in the gas, heating it slightly. Our next goal is to calculate the net energy exchange with the IGM as these photons scatter and eventually disappear. We begin with the redistribution function $R_{\text{II}}(x, x')$, which gives the probability that a photon absorbed at frequency x' is re-emitted at frequency x , assuming coherent scattering in the rest frame of the absorbing atom (Henyey, 1941; Hummer, 1962), thus ignoring recoil:

$$R_{\text{II}}(x, x') = \frac{1}{\pi^{3/2}} \int_{|\bar{x}-x|/2}^{\infty} du e^{-u^2} \left[\tan^{-1} \left(\frac{x+u}{a} \right) - \tan^{-1} \left(\frac{\bar{x}-u}{a} \right) \right], \quad (4.37)$$

where $\bar{x} = \max(x, x')$ and $\underline{x} = \min(x, x')$. This is normalized so that

$$\phi(x') = \int_{-\infty}^{\infty} dx R_{\text{II}}(x, x'). \quad (4.38)$$

In our case ($x' \gg 0$), the redistribution function is sharply peaked around x' . We thus write $x = x' + \Delta x$ and expand the inverse tangents to third order in $(u, \Delta x)$ about x' (note that, although u can be arbitrarily large, the exponential guarantees that only small values contribute to the integral). Equation (4.37) then becomes

$$R_{\text{II}}(x, x') \approx \frac{a}{\pi^{3/2} x'^2} \int_{|\Delta x|/2}^{\infty} du e^{-u^2} \left[(2u - \Delta x) + \frac{(\Delta x)^2 - 2u\Delta x}{x'} \right], \quad x > x', \quad (4.39)$$

with a similar expression when $x < x'$. We are interested in the mean energy loss in each scattering,

$$\langle \Delta x | x' \rangle = \phi^{-1}(x') \int_{-\infty}^{\infty} dx \Delta x R_{\text{II}}(x, x'), \quad (4.40)$$

$$= \frac{-2}{\pi^{1/2} x'} \int_0^{\infty} d\Delta x \Delta x^2 \int_{\Delta x/2}^{\infty} du e^{-u^2} (2u - \Delta x), \quad (4.41)$$

where we have used the normalization of R_{II} and substituted our expansion for the redistribution function in the second equality. The integrals are elementary and are most easily performed by switching the order of integration; the simple result is

$$\langle \Delta x | x' \rangle = -1/x'. \quad (4.42)$$

This is identical to the second term in equation (3) of Chuzhoy & Shapiro (2006b) in the appropriate limit.

As expected, photons tend to lose energy to the gas, but only slowly. Physically, we have assumed that the scattering is coherent in the rest frame of the atom. Thus, in the IGM frame, the gas tends to gain energy if the scattering atom travels away from the initial photon and to lose energy if the atom travels toward it. When scattering occurs blueward of resonance, the former have a slightly higher cross section for absorption because of the small blueshift imparted to them by their thermal velocity; thus the net effect is energy transfer to the gas. However, far out on the wings of the line, the difference in cross sections from this displacement is small, and the heating is weak.

In contrast, consider scattering with zero natural width (or in other words where the total optical depth is small, so that the initial scattering occurs in the Doppler core). In that case, R_{II} simplifies to $R_{\text{I}}(x, x') = \text{erfc}(\overline{|x|})/2$ (again assuming isotropic scattering in the rest frame), where $\overline{|x|} = \max(|x|, |x'|)$ (Unno, 1952). This has a flat core between $(-x', x')$, so the typical energy lost in the initial scattering is $\sim x'$. In this case, photons on the blue side are *only* scattered by atoms moving away from them (so that their frequency lines up with the infinitely sharp resonance), and in the lab frame the re-emitted photon typically shifts by a full Doppler width. Thus, in the few-scatterings limit, heating will be most efficient inside the line core. This is the case considered (for deuterium) by Chuzhoy & Shapiro (2006a).

Returning to the *Lyn* lines, where all the interactions occur in the wings, the net frequency shift in s scattering events is

$$\Delta x_{\text{tot}} \approx \frac{s(1+s)}{2x_1}, \quad (4.43)$$

$$\sim 0.033T_K^{1/2} \left[\frac{s_n(1+s_n)}{110} \right] \left(\frac{\nu_n}{\nu_\beta} \right)^4 \left(\frac{A_\beta}{A_n} \right) \left(\frac{\Gamma_\beta}{\Gamma_n} \right) \left(\frac{1+z_r}{20} \right)^{-3/2}. \quad (4.44)$$

Again, we have normalized to the values appropriate for $\text{Ly}\beta$ photons. Of course, we have assumed that the scatterings occur in the wings of the line. If the accumulated drift carries the photon toward line center, subsequent scattering will occur symmetrically and heating will be negligible. Thus we must have $\Delta x_{\text{tot}} \leq x_1$. This is marginally true for $\text{Ly}\epsilon$, which has $s_\epsilon = 5$ and a coefficient $0.77T_K^{1/2}$ when appropriate values are inserted into equation (4.44).

It is useful to compare this drift to that due to recoil itself, which we ignored by using R_{II} . This has $\Delta x_{\text{recoil}} = \eta$ per scattering. Thus

$$\frac{\Delta x_{\text{tot}}}{\Delta x_{\text{recoil}}} \sim 0.11T_K \left(\frac{1+s_n}{10} \right) \left(\frac{\nu_n}{\nu_\beta} \right)^4 \left(\frac{A_\beta}{A_n} \right) \left(\frac{\Gamma_\beta}{\Gamma_n} \right) \left(\frac{1+z_r}{20} \right)^{-3/2}; \quad (4.45)$$

the coefficient is 2.35 for $\text{Ly}\epsilon$ photons. Thus, at reasonably large temperatures, the frequency drift from repeated scattering overwhelms recoil. Pritchard & Furlanetto (2006) showed that recoil provides a negligibly small contribution at all temperatures. In practice, heating from *Lyn* scattering is never significant: only if $T_K \gg T_\gamma$ could it possibly matter, but in that case other, much stronger, heating agents must already be present.

As we have shown, Δx is largest (~ 1) when scattering occurs near the line core. One

example, considered by Chuzhoy & Shapiro (2006a), is the deuterium Ly β resonance, for which the optical depth is of order unity. Then the energy transfer is much more efficient, and the temperature dependence differs. In either case, $\Delta T \propto h\Delta\nu_D \langle \Delta x \rangle$ per scattering. When absorption is in the line center $\langle \Delta x \rangle \sim 1$; in the wings, we have seen that natural broadening controls the cross section and $\langle \Delta x \rangle \propto 1/x_1 \propto \Delta\nu_D$. So $\Delta T_{\text{core}} \propto T_K^{1/2}$ and $\Delta T_{\text{wing}} \propto T_K$. Deuterium Ly β turns out to be the most important transition (aside from hydrogen Ly α) in heat exchange and must be included in some circumstances. Of course, this energy is injected into the deuterium, rather than the hydrogen, to which it must be transferred by collisions. According to Chuzhoy & Shapiro (2006b), this is relatively inefficient. As a result, the deuterium temperature may become quite large ($T \sim 10^4$ K), where the wing approximation breaks down and the full Voigt profile must be used.

4.7 Discussion

We have examined both analytic and numeric solutions for the radiation field near the Ly α resonance and used them to compute the total scattering rate and the IGM heating (or cooling) rate. We showed that the approximate analytic solution of Grachev (1989) and Chuzhoy & Shapiro (2006b), in which scattering in the wings dominates, is accurate so long as $T_K \lesssim 1000$ K. At higher temperatures, thermal broadening becomes important. Fortunately, the scattering correction $S_\alpha \rightarrow 1$ at large temperatures. So the approximate fit presented by Chuzhoy & Shapiro (2006b) – our equation (4.20) – turns out to be reasonably accurate (to several percent) whenever $T_K \gtrsim 1$ K. For higher accuracy, equation (4.17) can be used.

We then used this analytic solution to examine the heating (or cooling) from the scattering near line center. For the case of photons that redshift toward the resonance, we obtained a fully analytic solution (in terms of Airy functions) under the approximation that all scattering occurs in the wings. The arguments of the Airy functions are typically small, so a power series expansion is illuminating; it shows that the heating rate per atom and per Hubble time is proportional to $T_K^{-1/3}(1+z)^{-2}$ when spin exchange can be neglected; this is an excellent approximation at $T_K \gtrsim 10$ K. In the case of photons injected at line center, we obtained a power series solution that converges reasonably rapidly. The lowest-order term is reasonably accurate for $T_K \gtrsim 100$ K, and in this regime the *cooling* rate per atom

Table 4.1: The quantity $(1 - S_\alpha)$ as a function of temperature T_K at $z = 20$ under several approximations (see text). Note that the fit in equation (4.20) was originally proposed by Chuzhoy & Shapiro (2006b).

Scenario	1 K	10 K	10^2 K	10^3 K	10^4 K
Normal	0.7798	0.3002	0.07516	0.01617	0.002484
Wing Approx	0.7795	0.3001	0.07527	0.01657	0.003175
Eq. (4.20)	0.8175	0.3068	0.07592	0.01689	0.003658
Detailed Balance	0.7798	0.3003	0.07528	0.01629	0.003137
Spin, $T_S = T_\gamma$	0.7121	0.2956	0.07546	0.01629	0.002511
Spin, $T_S = T_K$	0.8117	0.3034	0.07525	0.01618	0.002484

and per Hubble time is proportional to $T_K^{1/3}(1+z)^{-5/2}$ (again when spin exchange can be neglected). Photons injected in this way only heat the gas when $T_K \lesssim 10$ K.

Obviously, finding convenient and useful forms for S_α and the heating rates is a game of approximations. Because several different ones have been made in the literature, in Tables 4.1 and 4.2 we show the results for $(1 - S_\alpha)$ [essentially $\delta_J(0)$] and I_c in a variety of scenarios. The first row in each table gives our standard result, which includes the full Voigt profile and the detailed balance correction but ignores spin exchange and, of course, makes the Fokker-Planck approximation. The second row shows the results in the wing approximation of Grachev (1989) and Chuzhoy & Shapiro (2006b); it is excellent at low temperatures but begins to deviate by $\lesssim 10\%$ at higher temperatures where Voigt broadening is significant. The next row uses the approximate forms of equation (4.20) and (4.30); the latter is relatively poor at small temperatures because it is only a first-order expansion. Next, we show the effects of ignoring the detailed balance correction: it also makes no difference at small temperatures, but it matters at the $\sim 15\%$ level at $T = 10^4$ K because the line broadening becomes relatively significant compared to the rest frequency of the line. In the next two lines, we show how including spin exchange affects the results. Because this introduces a new variable (T_S), we show the two limiting cases where $T_S = T_\gamma$ and $T_S = T_K$, which bracket the possible effects. Recall that this introduces a new source of drift and diffusivity, with a magnitude $\sim \nu_{21}^2/\Delta\nu_D^2$ (Hirata, 2006; Chuzhoy & Shapiro, 2006b). Thus it can make a substantial difference when $T_K \ll 10$ K, but for most of the range of interest its effects are small.

Finally, nearly all of the existing literature – including our calculations – uses the Fokker-

Table 4.2: The quantity I_c as a function of temperature T_K at $z = 20$ under several approximations (see text).

Scenario	1 K	10 K	10^2 K	10^3 K	10^4 K
Normal	90.56	20.76	3.527	0.5346	0.06940
Wing Approx	90.54	20.72	3.506	0.5243	0.06558
First-Order (eq. 4.30)	170.2	24.98	3.662	0.5293	0.06569
Detailed Balance	90.56	20.76	3.533	0.5437	0.08346
Spin, $T_S = T_\gamma$	88.99	20.70	3.546	0.5383	0.06998
Spin, $T_S = T_K$	108.4	21.26	3.536	0.5347	0.06940

Planck approximation, which assumes that the background spectrum is constant over the typical frequency change per scattering; as we have seen, this requires slow changes over $\Delta x \sim 1$ at line center but is less restrictive in the wings. In general this is an excellent approximation; the background spectrum is in fact remarkably smooth (see Fig. 4.1). As emphasized by Hirata (2006), the worst-case scenario is for low-temperature gas (where the absorption trough is most sharply peaked) when spin diffusivity is included. Using a Monte Carlo model, he verified that the Fokker-Planck approximation for S_α is accurate to better than 3% at $T_K = 2$ and 10 K, at a variety of spin temperatures. There have been no explicit tests at higher temperatures, but (also as argued by Hirata 2006) it is expected to be even more accurate there. First of all, the absorption feature becomes less pronounced so that the assumption that $J(\nu)$ is constant over a frequency exchange is more accurate. (Also, $S_\alpha \approx 1$ anyway at high temperatures.) Second, spin effects become less significant, because the separate hyperfine lines become broadened into a single line. Meiksin (2006) also examined the Fokker-Planck approximation in the context of the heating rate, carrying the perturbative expansion to higher, post-diffusive order. Unfortunately, the resulting equations are not easy to manipulate or solve efficiently. In general, we expect the heating and cooling rates to be less sensitive to the Fokker-Planck assumptions than S_α because most of the absorption feature appears at large x , where the frequency change per scattering is much less than unity and the smoothness requirement is less severe. Nevertheless, continued exploration of its accuracy is an important unsettled question, and none of our solutions should be believed to better than a few percent.

A final ‘‘approximation’’ often made in the literature is to ignore Lyn photons. Their direct scattering contributes only $\lesssim 10^{-4}$ of the coupling (Pritchard & Furlanetto, 2006) or

$\lesssim 0.01$ of the heating (from the arguments in §4.6). However, $\sim 1/3$ of these photons cascade to Ly α (Hirata, 2006; Pritchard & Furlanetto, 2006), which can have significant effects. For example, with Population II stars, $\sim 6\%$ of the total Ly α background comes from cascades. Because injected photons can cause cooling, even this small flux can qualitatively change the implications for T_K (Chuzhoy & Shapiro, 2006a). They therefore cannot be ignored in this context.

As a final thought, it is useful to estimate the heating rate from Ly α scattering to gauge its importance relative to other processes. For simplicity, we will consider continuous injection (i.e., photons redshifting into the Ly α resonance). Inserting our lowest-order approximation for I_c (from eq. 4.30) into equation (4.25), we find

$$\frac{2}{3} \frac{\epsilon_\alpha}{H n_{\text{HI}} k_B T_K} \approx \frac{0.80}{T_K^{4/3}} \frac{x_\alpha}{S_\alpha} \left(\frac{10}{1+z} \right), \quad (4.46)$$

where the left hand side is the fractional temperature change per Hubble time. On the right hand side, we have rewritten $J(\nu)$ in terms of the 21 cm coupling efficiency x_α (see eq. 4.14). The 21 cm spin temperature departs from the CMB temperature when $x_\alpha \sim 1$, so this is a convenient gauge for the background flux at which heating first becomes observationally relevant. Clearly, Ly α heating is negligible at this point unless the initial temperature is also small. (Note that this approximation for I_c *overestimates* the heating at low temperatures, so the actual coefficient is even smaller than predicted by eq. 4.46.) Because, even without any heating, $T_K = 2.5$ K at $z = 10$ (Seager et al., 1999), Ly α scattering is unlikely to be significant in this context: it alone will not suppress a 21 cm absorption epoch (Chen & Miralda-Escudé, 2004).

Of course, the heating rate becomes much larger in strongly-coupled gas. However, in practice it is still probably negligible compared to other processes such as X-ray heating (Oh, 2001; Glover & Brand, 2003; Furlanetto, 2006). Following Furlanetto (2006), let us suppose that the X-ray emissivity traces the star formation rate (just like the Ly α emissivity). Then the ratio of X-ray and Ly α heating rates is

$$\frac{\epsilon_X}{\epsilon_\alpha} \sim 140 f_X T_K^{1/3} \left(\frac{f_{X,h}}{0.2} \frac{9690}{N_\alpha} \frac{1+z}{10} \right), \quad (4.47)$$

where we have again used our lowest-order approximation to I_c and T_K is in degrees Kelvin.

Here $f_{X,h}$ is the fraction of X-ray energy that is used to heat the gas (Shull & van Steenberg, 1985), N_α is the number of photons between Ly α and the Lyman-limit produced per baryon in stars (we have inserted the value appropriate for low-metallicity Pop II stars; Barkana & Loeb 2005b), and f_X is the assumed X-ray luminosity per unit star formation calibrated to the local value between 0.2 and 10 keV (Ranalli et al., 2003; Gilfanov et al., 2004). Clearly, Ly α heating is slow unless the X-ray emissivity is much smaller than its local value.

Acknowledgments

We thank G. Rybicki and M. Furlanetto for helpful discussions.

Chapter 5

21 cm fluctuations from inhomogeneous X-ray heating before reionization

Many models of early structure formation predict a period of heating immediately preceding reionization, when X-rays raise the gas temperature above that of the cosmic microwave background. These X-rays are often assumed to heat the intergalactic medium (IGM) uniformly, but in reality will heat the gas more strongly closer to the sources. We develop a framework for calculating fluctuations in the 21 cm brightness temperature that originate from this spatial variation in the heating rate. High-redshift sources are highly clustered, leading to significant gas temperature fluctuations (with fractional variations $\sim 40\%$, peaking on $k \sim 0.1 \text{ Mpc}^{-1}$ scales). This induces a distinctive peak-trough structure in the angle-averaged 21 cm power spectrum, which may be accessible to the proposed Square Kilometre Array. This signal reaches the $\sim 10 \text{ mK}$ level, and is stronger than that induced by $\text{Ly}\alpha$ flux fluctuations. As well as probing the thermal evolution of the IGM before reionization, this 21 cm signal contains information about the spectra of the first X-ray sources. Finally, we consider disentangling temperature, density, and $\text{Ly}\alpha$ flux fluctuations as functions of redshift.

Originally published as Pritchard and Furlanetto, MNRAS, **376**, 1680 (2007).

5.1 Introduction

The formation of the first luminous objects ends the cosmic “dark ages” and begins a period of heating and ionization of the intergalactic medium (IGM). The global thermodynamic history of this epoch, which culminates in reionization, depends upon many poorly constrained processes such as star formation, radiative feedback, and the growth of HII regions (Barkana & Loeb, 2001). Currently, the best constraints on the ionization history come

from observations of the Gunn-Peterson trough in quasar absorption lines (Gunn & Peterson, 1965) and in WMAP observations of the optical depth to recombination (Spergel et al., 2006). Current observations of the temperature evolution of the IGM are similarly limited. At low redshift, observations of the Ly α forest place constraints on the temperature of the IGM after reionization (Schaye et al., 2000; McDonald et al., 2001; Zaldarriaga et al., 2001; Theuns et al., 2002; Hui & Haiman, 2003). Unfortunately, photoionization during reionization causes a large temperature increase that essentially erases information about the preceding period. At high redshift, it is assumed that the gas cools adiabatically after thermal decoupling from the CMB at $z \approx 150$, when Compton scattering becomes inefficient (Peebles, 1993). The intermediate regime, where the first sources have “switched on”, is poorly constrained. Once collapsed structures form many different heating mechanisms are possible, e.g., shock heating (Furlanetto & Loeb, 2004), resonant scattering of Ly α photons (Madau, Meiksin & Rees, 1997; Chen & Miralda-Escudé, 2004; Chuzhoy & Shapiro, 2006a; Meiksin, 2006; Rybicki, 2006; Furlanetto & Pritchard, 2006), and X-ray heating (Ostriker & Gnedin, 1996; Oh, 2001; Venkatesan, Giroux & Shull, 2001; Ricotti & Ostriker, 2004). Determining the thermal history and identifying the important heating mechanisms requires new observations.

Future telescopes such as the *James Webb Space Telescope* hope to image high-redshift sources directly. However, seeing the sources is not the same as seeing the heating and ionization they cause in the IGM. The most promising technique for probing the thermal history of the IGM before reionization is via observation of the 21 cm hyperfine transition of neutral hydrogen (Furlanetto et al., 2006, and references therein). This line may be seen in absorption against the CMB, when the spin temperature T_S is less than the CMB temperature T_γ , or in emission, when $T_S > T_\gamma$. Three prototype low-frequency interferometers (LOFAR¹, MWA², and PAST³) are under construction and should be capable of observing the redshifted 21 cm signal from gas at redshifts $z \lesssim 12$, with the proposed Square Kilometre Array⁴ (SKA) capable of probing even higher redshifts. A great deal of theoretical work has now been done in calculating the 21 cm signal from fluctuations in density δ (Loeb & Zaldarriaga, 2004), the Ly α flux J_α (Barkana & Loeb, 2005b; Pritchard & Furlanetto,

¹See <http://www.lofar.org/>.

²See <http://web.haystack.mit.edu/arrays/MWA/>.

³See Pen, Wu & Peterson (2005).

⁴See Carilli & Rawlings (2004).

2006), and the neutral fraction (Zaldarriaga, Furlanetto & Hernquist, 2004; Furlanetto, Zaldarriaga & Hernquist, 2004). Fluctuations in the 21 cm brightness temperature T_b also occur because of fluctuations in the gas kinetic temperature T_K , but this has not yet been explored.

In this paper, we explore the effect of inhomogeneous X-ray heating by the first luminous sources on the 21 cm signal using analytic techniques. We first build a model for the global thermal history of the IGM following Furlanetto (2006). In this model, we assume that a population of X-ray sources resulting from the remnants of the first stars is responsible for heating the IGM (Ostriker & Gnedin, 1996; Oh, 2001; Venkatesan et al., 2001; Ricotti & Ostriker, 2004). X-ray heating is dominated by soft X-rays ($E \lesssim 2$ keV), as harder X-rays have a mean free path comparable with the Hubble scale. These long mean free paths have often motivated the simplifying assumption that X-rays heat the IGM uniformly. In fact, clustering of the sources and the $1/r^2$ decrease of flux with distance combine to produce significantly inhomogeneous heating. We develop a formalism, based upon that of Barkana & Loeb (2005b), for calculating the temperature fluctuations that are sourced by these inhomogeneities. We use this to explore features in the 21 cm power spectrum that constrain the evolution of T_K . This calculation also motivates a consideration of the possibility of using 21 cm measurements to constrain the X-ray emission spectrum of the first sources.

Simulations of the early universe have yet to address the spectrum of temperature fluctuations in the period before reionization. Previous analytic consideration of fluctuations in T_K has focused on the period following recombination but before sources form (Barkana & Loeb, 2005c; Naoz & Barkana, 2005). Temperature fluctuations induced by the first sources have not previously been considered in detail.

The 21 cm signal can be thought of as a tool for probing various radiation backgrounds. Gas temperature fluctuations probe the X-ray background, neutral fraction fluctuations probe the ionizing UV background, and Ly α fluctuations probe the non-ionizing UV background. While the focus of this paper is X-ray heating of the IGM, in practice, the different sources of 21 cm fluctuation are not cleanly separated. In order to properly establish context, we briefly re-examine the signal from fluctuations in the Ly α flux, incorporating Ly α production by X-ray excitation of HI (Chen & Miralda-Escude, 2006; Chuzhoy, Alvarez & Shapiro, 2006), and determine whether this contains extra useful information for con-

straining the spectral properties of the X-ray sources. Finally, we explore the feasibility of separating information on the temperature and Ly α flux fluctuations with the 21 cm signal.

The layout of this paper is as follows. We begin by setting out the physics of the 21 cm signal in §5.2. Calculating this requires a model for the global history of the IGM, which we outline in §5.3. Having established the mean history, in §5.4 we describe our framework for calculating fluctuations in T_K , J_α , and the neutral fraction. This is used to calculate the power spectrum for fluctuations in T_K in §5.5. We then calculate the 21 cm signal in §5.6, exploring the redshift evolution and dependence on the X-ray source spectrum and luminosity. Finally, we discuss the possibility of observationally detecting and separating these signals in §5.7 before concluding in §5.8. Throughout this paper, we assume a cosmology with $\Omega_m = 0.26$, $\Omega_\Lambda = 0.74$, $\Omega_b = 0.044$, $H = 100h \text{ km s}^{-1} \text{ Mpc}^{-1}$ (with $h = 0.74$), $n_S = 0.95$, and $\sigma_8 = 0.8$, consistent with the latest measurements (Spergel et al., 2006), although we have increased σ_8 above the best-fit WMAP value to improve agreement with weak-lensing data.

5.2 21 cm signal

We begin by briefly summarising the physics of the 21 cm signal and refer the interested reader to Furlanetto, Oh & Briggs (2006) for further information. The 21 cm line of the hydrogen atom results from hyperfine splitting of the $1S$ ground state due to the interaction of the magnetic moments of the proton and the electron. The HI spin temperature T_S is defined via the number density of hydrogen atoms in the $1S$ singlet and triplet levels, n_0 and n_1 respectively, $n_1/n_0 = (g_1/g_0) \exp(-T_\star/T_S)$, where $(g_1/g_0) = 3$ is the ratio of the spin degeneracy factors of the two levels, and $T_\star \equiv hc/k\lambda_{21\text{cm}} = 0.0628 \text{ K}$. The optical depth of this transition is small at all relevant redshifts, so the brightness temperature of the CMB is

$$T_b = 27x_{\text{HI}}(1 + \delta_b) \times \left(\frac{\Omega_b h^2}{0.023} \right) \left(\frac{0.15}{\Omega_m h^2} \frac{1+z}{10} \right)^{1/2} \left(\frac{T_S - T_\gamma}{T_S} \right) \text{ mK}, \quad (5.1)$$

Here x_{HI} is the neutral fraction of hydrogen and δ_b is the fractional overdensity in baryons. The spin temperature is given by

$$T_S^{-1} = \frac{T_\gamma^{-1} + x_\alpha T_\alpha^{-1} + x_c T_K^{-1}}{1 + x_\alpha + x_c}, \quad (5.2)$$

where T_α is the colour temperature of the Ly α radiation field at the Ly α frequency and is closely coupled to T_K by recoil during repeated scattering. The spin temperature becomes strongly coupled to the gas temperature when $x_{\text{tot}} \equiv x_c + x_\alpha \gtrsim 1$.

The collisional coupling coefficient is given by

$$x_c = \frac{4T_\star}{3A_{10}T_\gamma} [\kappa_{1-0}^{HH}(T_k)n_H + \kappa_{1-0}^{eH}(T_k)n_e], \quad (5.3)$$

where $A_{10} = 2.85 \times 10^{-15} \text{ s}^{-1}$ is the spontaneous emission coefficient, κ_{1-0}^{HH} is tabulated as a function of T_k (Allison & Dalgarno, 1969; Zygelman, 2005) and κ_{1-0}^{eH} is taken from Furlanetto & Furlanetto (2006). For a more detailed analysis of the collisional coupling, see Hirata & Sigurdson (2006).

The Wouthysen-Field effect (Wouthuysen, 1952; Field, 1958) coupling is given by

$$x_\alpha = \frac{16\pi^2 T_\star e^2 f_\alpha}{27 A_{10} T_\gamma m_e c} S_\alpha J_\alpha, \quad (5.4)$$

where $f_\alpha = 0.4162$ is the oscillator strength of the Ly α transition. S_α is a correction factor of order unity, which describes the detailed structure of the photon distribution in the neighbourhood of the Ly α resonance (Chen & Miralda-Escudé, 2004; Hirata, 2006; Chuzhoy & Shapiro, 2006a; Furlanetto & Pritchard, 2006). We make use of the approximation for S_α outlined in Furlanetto & Pritchard (2006). For the models considered in this paper, Ly α coupling dominates over collisional coupling.

Fluctuations in the 21 cm signal may be expanded (Furlanetto et al., 2006)

$$\delta_{T_b} = \beta\delta + \beta_x\delta_x + \beta_\alpha\delta_\alpha + \beta_T\delta_T - \delta_{\partial v}, \quad (5.5)$$

where each δ_i describes the fractional variation in the quantity i : δ_α for fluctuations in the Ly α coupling coefficient, δ_x for the neutral fraction, δ_T for T_K , and $\delta_{\partial v}$ for the line-of-sight

peculiar velocity gradient. The expansion coefficients are given by

$$\begin{aligned}
\beta &= 1 + \frac{x_c}{x_{\text{tot}}(1 + x_{\text{tot}})}, \\
\beta_x &= 1 + \frac{x_c^{HH} - x_c^{eH}}{x_{\text{tot}}(1 + x_{\text{tot}})}, \\
\beta_\alpha &= \frac{x_\alpha}{x_{\text{tot}}(1 + x_{\text{tot}})}, \\
\beta_T &= \frac{T_\gamma}{T_K - T_\gamma} + \frac{1}{x_{\text{tot}}(1 + x_{\text{tot}})} \left(x_c^{eH} \frac{d \log \kappa_{10}^{eH}}{d \log T_K} + x_c^{HH} \frac{d \log \kappa_{10}^{HH}}{d \log T_K} \right).
\end{aligned} \tag{5.6}$$

In this, we assume that baryons trace the density field exactly so that $\delta_b = \delta$. All of these quantities are positive, with the exception of β_T , whose sign is determined by $(T_K - T_\gamma)$. The apparent divergence in β_T when $T_K = T_\gamma$ is an artefact of expanding the fractional brightness temperature about a point where the mean brightness temperature $\bar{T}_b = 0$. The physical quantity $\bar{T}_b \beta_T$ is always well behaved.

Noting that in Fourier space $\delta_{\partial v} = -\mu^2 \delta$ (Bharadwaj & Ali, 2004), where μ is the angle between the line of sight and the wavevector \mathbf{k} of the Fourier mode, we may use equation (5.5) to form the power spectrum (Barkana & Loeb, 2005a)

$$P_{T_b}(k, \mu) = P_{\mu^0}(k) + \mu^2 P_{\mu^2}(k) + \mu^4 P_{\mu^4}(k). \tag{5.7}$$

In theory, high precision measurements of the 3D power spectrum will allow the separation of these terms by their angular dependence. However, it is unclear whether the first generation of 21 cm experiments will be able to achieve the high signal-to-noise required for this separation (McQuinn et al., 2005). Instead, they will measure the angle averaged quantity

$$\bar{P}_{T_b}(k) = P_{\mu^0}(k) + P_{\mu^2}(k)/3 + P_{\mu^4}(k)/5. \tag{5.8}$$

In presenting our results, we will concentrate on $P_{\mu^2}(k)$, which most cleanly separates out the different types of fluctuation, and $\bar{P}_{T_b}(k)$, which is easiest to observe. We will typically plot the power per logarithmic interval $\Delta = [k^3 P(k)/2\pi^2]^{1/2}$.

5.3 Global history

5.3.1 Outline

We may express T_b as a function of four variables $T_b = T_b(T_K, x_i, J_\alpha, n_H)$. In calculating the 21 cm signal, we require a model for the global evolution of and fluctuations in these quantities. We will follow the basic formalism of Furlanetto (2006), but first let us consider the main events in likely chronological order. This determines redshift intervals where the signal is dominated by fluctuations in the different quantities.

$z \gtrsim 200$: After recombination, Compton scattering maintains thermal coupling of the gas to the CMB, setting $T_K = T_\gamma$ so that we expect $\bar{T}_b = 0$.

$40 \lesssim z \lesssim 200$: In this regime, adiabatic cooling means $T_K < T_\gamma$ and collisional coupling sets $T_S < T_\gamma$, leading to $\bar{T}_b < 0$ and a possible absorption signal. At this time, T_b fluctuations are sourced by density fluctuations, potentially allowing cosmology to be probed (Loeb & Zaldarriaga, 2004; Hirata & Sigurdson, 2006).

$z_* \lesssim z \lesssim 40$: As the expansion continues, decreasing the gas density, collisional coupling becomes ineffective, absorption of CMB photons sets $T_S = T_\gamma$, and there is no detectable 21 cm signal.

$z_\alpha \lesssim z \lesssim z_*$: Once the first sources switch on at z_* , they emit both Ly α photons and X-rays. In general, the emissivity required for Ly α coupling is significantly less than that for heating T_K above T_γ . Thus, in the simplest models, we expect the redshift z_α , where Ly α coupling saturates $x_\alpha \gg 1$, to be greater than z_h , where $\bar{T}_K = T_\gamma$. In this regime, $T_S \sim T_k < T_\gamma$ and there is an absorption signal. Fluctuations are dominated by density fluctuations and variation in the Ly α flux (Barkana & Loeb, 2005b; Pritchard & Furlanetto, 2006; Chen & Miralda-Escude, 2006).

$z_h \lesssim z \lesssim z_\alpha$: After Ly α coupling saturates, fluctuations in the Ly α flux no longer affect the 21 cm signal. By this point, heating becomes significant and gas temperature fluctuations source T_b fluctuations. While T_K remains below T_γ we see a 21 cm signal in absorption, but as T_K approaches T_γ hotter regions may begin to be seen in emission.

$z_T \lesssim z \lesssim z_h$: After the heating transition, $T_K > T_\gamma$ and we expect to see a 21 cm signal in emission. The 21 cm brightness temperature is not yet saturated, which occurs at z_T , when $T_S \sim T_K \gg T_\gamma$. By this time, the ionization fraction has likely risen above the percent level. Brightness temperature fluctuations are sourced by a mixture of fluctuations

in ionization, density and gas temperature.

$z_r \lesssim z \lesssim z_T$: Continued heating drives $T_K \gg T_\gamma$ at z_T and temperature fluctuations become unimportant. $T_S \sim T_K \gg T_\gamma$ and the dependence on T_S may be neglected in equation (5.1), which greatly simplifies analysis of the 21 cm power spectrum (Santos & Cooray, 2006). By this point, the filling fraction of HII regions probably becomes significant and ionization fluctuations begin to dominate the 21 cm signal (Furlanetto et al., 2004).

$z \lesssim z_r$: After reionization, any remaining 21 cm signal originates from overdense regions of collapsed neutral hydrogen.

Most of these epochs are not sharply defined, so there should be considerable overlap between them. This seems the most likely sequence of events, although there is considerable uncertainty in the ordering of z_α and z_h . Nusser (2005) explores the possibility that $z_h > z_\alpha$, so that X-ray preheating allows collisional coupling to be important before the Ly α flux becomes significant. Simulations of the very first mini-quasar (Kuhlen & Madau, 2005; Kuhlen, Madau & Montgomery, 2006) also probe this regime and show that the first luminous X-ray sources can have a great impact on their surrounding environment. We note that these authors ignored Ly α coupling, and that an X-ray background may generate significant Ly α photons (Chen & Miralda-Escude, 2006), as we discuss in §5.3.3.

In this paper, we will concentrate on the period after z_* , when luminous sources “switch on”, but before the IGM has been heated to temperatures $T_K \gg T_\gamma$ (our z_T). In this regime, Ly α coupling dominates and the 21 cm signal is seen in absorption at high z but in emission at lower z . We shall explore this transition in more detail below. One of our key observables for 21 cm observations is the sign of β_T , which indicates whether $T_K > T_\gamma$ (provided that collisional coupling can be neglected).

5.3.2 Heating and ionization

Having set the broad context, let us tighten our discussion with a concrete model for the evolution of the IGM; in this we follow Furlanetto (2006). We will distinguish between the ionization fraction x_i , relating to the volume filled by the highly ionized HII regions that are located around clusters of sources, and the free electron fraction x_e of the largely neutral gas outside these HII regions. The former is important for determining when reionization occurs, while the latter governs X-ray heating in the bulk of the IGM. We note that the volume filling fraction of the HII regions is well approximated by x_i , which we will use to

calculate volume averaged quantities. We further distinguish between T_K , the temperature of the IGM outside the HII regions, and the temperature of these photoionized regions $T_{\text{HII}} \approx 10^4$ K. At high z , these regions are small and will not have a significant effect; while at low z , where reionization is well advanced, these HII regions will dominate and invalidate our formalism.

We begin by writing down equations for the evolution of T_K , x_i , and x_e

$$\frac{dT_K}{dt} = \frac{2T_K}{3n} \frac{dn}{dt} + \frac{2}{3k_B} \sum_j \frac{\epsilon_j}{n}, \quad (5.9)$$

$$\frac{dx_i}{dt} = (1 - x_e)\Lambda_i - \alpha_A C x_i^2 n_H, \quad (5.10)$$

$$\frac{dx_e}{dt} = (1 - x_e)\Lambda_e - \alpha_A C x_e^2 n_H, \quad (5.11)$$

where ϵ_j is the heating rate per unit volume, and we sum over all possible sources of heating/cooling j . We define Λ_i to be the rate of production of ionizing photons per unit time per baryon applied to HII regions, Λ_e is the equivalent quantity in the bulk of the IGM, $\alpha_A = 4.2 \times 10^{-13} \text{cm}^3 \text{s}^{-1}$ is the case-A recombination coefficient⁵ at $T = 10^4$ K, and $C \equiv \langle n_e^2 \rangle / \langle n_e \rangle^2$ is the clumping factor. We model the clumping factor using $C = 2$; this value for C reproduces the qualitative form of the histories in Furlanetto (2006) and ensures reionization at $z \gtrsim 6$. This approximation is appropriate only while x_i is small, and will fail towards the end of reionization, when clumping becomes important in determining the effect of recombinations (Miralda-Escudé, Haehnelt & Rees, 2000).

In modelling the growth of HII regions, we take

$$\Lambda_i = \zeta(z) \frac{df_{\text{coll}}}{dt}, \quad (5.12)$$

where $f_{\text{coll}}(z)$ is the fraction of gas inside collapsed objects at z and the ionization efficiency parameter ζ is given by

$$\zeta = A_{\text{He}} f_{\star} f_{\text{esc}} N_{\text{ion}}, \quad (5.13)$$

with N_{ion} the number of ionizing photons per baryon produced in stars, f_{\star} the fraction of baryons converted into stars, f_{esc} the fraction of ionizing photons that escape the host halo,

⁵Note that we use the case-A value, which amounts to assuming that ionizing photons are absorbed inside dense, neutral systems (Miralda-Escudé et al., 2000)

and A_{He} a correction factor for the presence of Helium. This model for x_i is motivated by a picture of HII regions expanding into neutral hydrogen (Barkana & Loeb, 2001). In calculating f_{coll} , we use the Press & Schechter (1974) mass function dn/dm and determine a minimum mass m_{min} for collapse by requiring the virial temperature $T_{\text{vir}} \geq 10^4 \text{ K}$, appropriate for cooling by atomic hydrogen. Decreasing this minimum mass, say to that of molecular cooling, will allow star formation to occur at earlier times shifting the features that we describe in redshift. We note that $x_e \ll 1$ at all redshifts under consideration, as once the free electron fraction reaches a few percent further X-ray energy is deposited primarily as heat, not further ionization.

To integrate equation (5.9), we must specify which heating mechanisms are important. Furlanetto (2006) considers several heating mechanisms including shock heating (Furlanetto & Loeb, 2004) and resonant scattering of Ly α photons (Madau et al., 1997; Chen & Miralda-Escudé, 2004; Chuzhoy & Shapiro, 2006a; Furlanetto & Pritchard, 2006). We shall neglect these contributions to heating of the IGM, focusing instead on the dominant mechanisms of Compton heating and X-ray heating. While shock heating dominates the thermal balance at low z , during the epoch we are considering it, probably, heats the gas only slightly before X-ray heating dominates.

Compton heating serves to couple T_K to T_γ at redshifts $z \gtrsim 150$, but becomes ineffective below that redshift. In our context, it serves to set the initial conditions before star formation begins. The heating rate per particle for Compton heating is given by

$$\frac{2 \epsilon_{\text{compton}}}{3 k_B n} = \frac{x_e}{1 + f_{\text{He}} + x_e} \frac{T_\gamma - T_K}{t_\gamma} \frac{u_\gamma}{\bar{u}_\gamma} (1 + z)^4, \quad (5.14)$$

where f_{He} is the helium fraction (by number), u_γ is the energy density of the CMB, $\sigma_T = 6.65 \times 10^{-25} \text{ cm}^2$ is the Thomson cross-section, and we define

$$t_\gamma^{-1} = \frac{8 \bar{u}_\gamma \sigma_T}{3 m_e c} = 8.55 \times 10^{-13} \text{ yr}^{-1}. \quad (5.15)$$

X-rays heat the gas primarily through photo-ionization of HI and HeI; this generates energetic photo-electrons, which dissipate their energy into heating, secondary ionizations, and atomic excitation. With this in mind, we calculate the total rate of energy deposition

per unit volume as

$$\epsilon_X = 4\pi n_i \int d\nu \sigma_{\nu,i} J_\nu (h\nu - h\nu_{\text{th}}), \quad (5.16)$$

where we sum over the species $i = \text{HI}$, HeI , and HeII ; n_i is the number density of species i ; $h\nu_{\text{th}} = E_{\text{th}}$ is the threshold energy for ionization; $\sigma_{\nu,i}$ is the cross-section for photoionization; and J_ν is the number flux of photons of frequency ν . We may divide this energy into heating, ionization, and excitation by inserting the factor $f_i(\nu)$, defined as the fraction of energy converted into form i at a specific frequency. The relevant division of the X-ray energy depends on x_e and is calculated using the fitting formulae of Shull & van Steenberg (1985). The $f_i(\nu)$ are approximately independent of ν for $h\nu \gtrsim 100$ eV, so that the ionization rate is related to the heating rate by a factor $f_{\text{ion}}/(f_{\text{heat}}E_{\text{th}})$. The X-ray number flux is found from

$$\begin{aligned} J_X(z) &= \int_{\nu_{\text{th}}}^{\infty} d\nu J_X(\nu, z), \\ &= \int_{\nu_{\text{th}}}^{\infty} d\nu \int_z^{z^*} dz' \frac{(1+z)^2}{4\pi} \frac{c}{H(z')} \hat{\epsilon}_X(\nu', z') e^{-\tau}, \end{aligned} \quad (5.17)$$

where $\hat{\epsilon}_X(\nu, z)$ is the comoving photon emissivity for X-ray sources, and ν' is the emission frequency at z' corresponding to an X-ray frequency ν at z

$$\nu' = \nu \frac{(1+z')}{(1+z)}. \quad (5.18)$$

The optical depth is given by

$$\tau(\nu, z, z') = \int_z^{z'} \frac{dl}{dz''} dz'' [n_{\text{HI}}\sigma_{\text{HI}}(\nu'') + n_{\text{HeI}}\sigma_{\text{HeI}}(\nu'') + n_{\text{HeII}}\sigma_{\text{HeII}}(\nu'')], \quad (5.19)$$

where we calculate the cross-sections using the fits of Verner et al. (1996). Care must be taken here, as the cross-sections have a strong frequency dependence and the X-ray frequency can redshift considerably between emission and absorption. In practice, the abundance of HeII is negligible and may be neglected.

X-ray heating is often portrayed as uniform, as the X-ray photons possess long mean free paths. The comoving mean free path of an X-ray with energy E is (Furlanetto et al.,

2006)

$$\lambda_X \approx 4.9 \bar{x}_{\text{HI}}^{-1/3} \left(\frac{1+z}{15} \right)^{-2} \left(\frac{E}{300 \text{ eV}} \right)^3 \text{ Mpc}. \quad (5.20)$$

Thus, the Universe will be optically thick, over a Hubble length, to all photons with energy below $E \sim 2[(1+z)/15]^{1/2} \bar{x}_{\text{HI}}^{1/3}$ keV. The E^{-3} dependence of the cross-section means that heating is dominated by soft X-rays, which do fluctuate on small scales. In addition, though, there will be a uniform component to the heating from harder X-rays.

We consider three possible sources of X-rays: starburst galaxies, supernova remnants (SNR), and miniquasars (Oh, 2001; Glover & Brand, 2003; Furlanetto, 2006). The incidences of starbursts and supernova remnants are likely to be tied to the global star formation rate (Glover & Brand, 2003). For simplicity, we will assume that miniquasars similarly track the star formation rate (SFR). In reality, of course, their evolution could be considerably more complex (Madau et al., 2004). We characterise these sources by an emissivity per unit (comoving) volume per unit frequency

$$\hat{\epsilon}_X(z, \nu) = \hat{\epsilon}_X(\nu) \left(\frac{\text{SFRD}}{\text{M}_\odot \text{ yr}^{-1} \text{ Mpc}^{-3}} \right), \quad (5.21)$$

where SFRD is the star formation rate density, and the spectral distribution function is a power law with index α_S

$$\hat{\epsilon}_X(\nu) = \frac{L_0}{h\nu_0} \left(\frac{\nu}{\nu_0} \right)^{-\alpha_S-1}, \quad (5.22)$$

and the pivot energy $h\nu_0 = 1$ keV. We assume emission within the band 0.1 – 30 keV, and set $L_0 = 3.4 \times 10^{40} f_X \text{ erg s}^{-1} \text{ Mpc}^{-3}$, where f_X is a highly uncertain constant factor. This normalisation is chosen so that, with $f_X = 1$, the total X-ray luminosity per unit SFR is consistent with that observed in starburst galaxies in the present epoch (see Furlanetto, 2006, for further details). Extrapolating observations from the present day to high redshift is fraught with uncertainty, and we note that this normalisation is very uncertain. The total X-ray luminosity at high redshift is constrained by observations of the present day soft X-ray back ground, which rules out complete reionization by X-rays, but allows considerable latitude for heating (Dijkstra, Haiman & Loeb, 2004). Similarly, there is significant uncertainty in the spectra of these objects. We choose $\alpha_S = 1.5$ for starbursts, $\alpha_S = 1.0$ for SNR, and $\alpha_S = 0.5$ for miniquasars (Madau et al., 2004). These span the reasonable spectral dependence of possible X-ray sources.

As in equation (5.12), we model the star formation rate as tracking the collapse of matter, so that we may write the star formation rate per (comoving) unit volume

$$\text{SFRD} = \bar{\rho}_b^0(z) f_* \frac{d}{dt} f_{\text{coll}}(z). \quad (5.23)$$

where $\bar{\rho}_b^0$ is the cosmic mean baryon density today. This formalism is appropriate for $z \gtrsim 10$, as at later times star formation as a result of mergers becomes important.

5.3.3 Ly α flux

Finally, we must describe the evolution of the Ly α flux. This is produced by stellar emission ($J_{\alpha,\star}$) and by X-ray excitation of HI ($J_{\alpha,X}$). Photons emitted by stars, between Ly α and the Lyman limit, will redshift until they enter a Lyman series resonance. Subsequently, they may generate Ly α photons, as discussed in Pritchard & Furlanetto (2006) and Hirata (2006). The Ly α flux from stars $J_{\alpha,\star}$ arises from a sum over the Ly n levels, with the maximum n that contributes $n_{\text{max}} \approx 23$ determined by the size of the HII region of a typical (isolated) galaxy (see Barkana & Loeb, 2005b, for details). The average Ly α background is then

$$\begin{aligned} J_{\alpha,\star}(z) &= \sum_{n=2}^{n_{\text{max}}} J_{\alpha}^{(n)}(z), \\ &= \sum_{n=2}^{n_{\text{max}}} f_{\text{recycle}}(n) \int_z^{z_{\text{max}}(n)} dz' \frac{(1+z)^2}{4\pi} \frac{c}{H(z')} \hat{\epsilon}_{\star}(\nu'_n, z'), \end{aligned} \quad (5.24)$$

where $z_{\text{max}}(n)$ is the maximum redshift from which emitted photons will redshift into the level n Lyman resonance, ν'_n is the emission frequency at z' corresponding to absorption by the level n at z , $f_{\text{recycle}}(n)$ is the probability of producing a Ly α photon by cascade from level n , and $\hat{\epsilon}_{\star}(\nu, z)$ is the comoving photon emissivity for stellar sources. We calculate $\hat{\epsilon}_{\star}(\nu, z)$ in the same way as for X-rays (eq. 21), and define $\hat{\epsilon}_{\star}(\nu)$ to be the spectral distribution function of the stellar sources. We consider models with Pop. II and very massive Pop. III stars. In each case, we take $\hat{\epsilon}_{\star}(\nu)$ to be a broken power law with one index describing emission between Ly α and Ly β , and a second describing emission between Ly β and the Lyman limit (see Pritchard & Furlanetto, 2006, for details).

Photoionization of HI or HeI by X-rays may also lead to the production of Ly α photons. In this case, some of the primary photo-electron's energy ends up in excitations of HI (Shull

& van Steenberg, 1985), which on relaxation may generate Ly α photons (Madau et al., 1997; Chen & Miralda-Escude, 2006; Chuzhoy et al., 2006). This Ly α flux $J_{\alpha,X}$ may be related to the X-ray heating rate as follows. The rate at which X-ray energy is converted into Ly α photons is given by

$$\epsilon_{X,\alpha} = \epsilon_{X,\text{heat}} \frac{f_{\text{ex}}}{f_{\text{heat}}} p_{\alpha}, \quad (5.25)$$

where f_{ex} and f_{heat} are the fraction of X-ray energy going into excitation and heating respectively, and p_{α} is the fraction of excitation energy that goes into Ly α photons. We then find the Ly α flux by assuming that this injection rate is balanced by photons redshifting out of the Ly α resonance, so

$$J_{\alpha,X} = \frac{c}{4\pi} \frac{\epsilon_{X,\alpha}}{h\nu_{\alpha}} \frac{1}{H\nu_{\alpha}}. \quad (5.26)$$

Shull & van Steenberg (1985) calculated f_{ex} and f_{heat} , but their Monte Carlo method, gives only a little insight into the value of p_{α} . Although excitations to the 2P level will always generate Ly α photons, only some fraction of excitations to other levels will lead to Ly α generating cascades. The rest will end with two photon decay from the 2S level. Shull & van Steenberg (1985) considered a simplified atomic system, in which collisional excitations to $n \geq 3$ levels were incorporated by multiplying the excitation cross-section to the $n = 2$ level by a factor of 1.35 (Shull, 1979). Thus, we might expect of order $0.35/1.35 \sim 0.26$ of collisional excitations to end at an $n \geq 3$ level.

We estimate p_{α} by calculating the probability that a secondary electron of energy E_{sec} will excite HI from the ground state to the level nL, using the collisional cross-sections⁶ of Bray et al. (2002), and then applying the probability that the resulting cascade will produce a Ly α photon, taken from Pritchard & Furlanetto (2006) and Hirata (2006). The iterative procedure of Pritchard & Furlanetto (2006) gives the probability of producing a Ly α photon by cascade from the level nL as: (0, 1) for (2S, 2P), (1, 0, 1) for (3S, 3P, 3D), and (0.584, 0.261, 0.746, 1) for (4S, 4P, 4D, 4F).

Summing over atomic levels $n \leq 4$, we obtain $p_{\alpha} = 0.79$ for $E_{\text{sec}} = 30$ eV. The contribution from $n > 4$ levels is small as the collisional cross-sections drop off rapidly as n increases. The exact result depends upon the energy distribution of the secondary electrons, which in turn depends upon the spectrum of ionizing X-rays. Our chosen value for E_{sec} corresponds to the mean electron energy (obtained using the distribution of Shull, 1979) produced by

⁶Taken from <http://atom.murdoch.edu.au/CCC-WWW/index.html>.

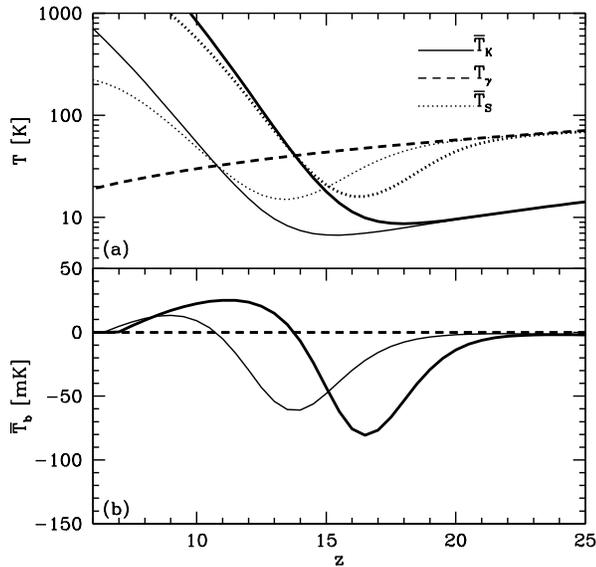


Figure 5.1: Mean IGM thermal history for model A (thick curves) and B (thin curves). (a): \bar{T}_K (solid curve), T_γ (dashed curve), and \bar{T}_S (dotted curve). (b): Volume averaged \bar{T}_b (solid curve). The zero line is indicated by a dashed horizontal line. Note that this is the thermal history outside of the ionized HII regions.

X-rays of energy 1.7 keV, which is the mean X-ray energy from a source with spectral index $\alpha = 1.5$ over the band 0.1–30 keV. Calculating p_α exactly requires an update of the Shull & van Steenberg (1985) calculation, but, by considering different values for E_{sec} , we conclude that it should differ from $p_\alpha = 0.79$ by less than 10%.

5.3.4 Model histories

Having outlined the various elements of our global history, we will restrict ourselves to considering two models. These will be A) Pop. II stars + starburst galaxies and B) Pop. III + starburst galaxies. Of course, these are only two of an infinite set of possibilities, but they serve to illustrate the effect of different Ly α and X-ray luminosities on the signal. We use parameters corresponding to Pop. II ($f_{\text{esc}} = 0.1$, $f_\star = 0.1$, $N_{\text{ion}} = 4000$) and very massive Pop. III ($f_{\text{esc}} = 0.1$, $f_\star = 0.01$, $N_{\text{ion}} = 30000$) stars (Furlanetto, 2006), although we note that these values are highly uncertain. We take $f_X = 1$ in both models, to allow straightforward comparison between the two models. The amplitude of the X-ray background is extremely uncertain, so that f_X is almost unconstrained, and we defer

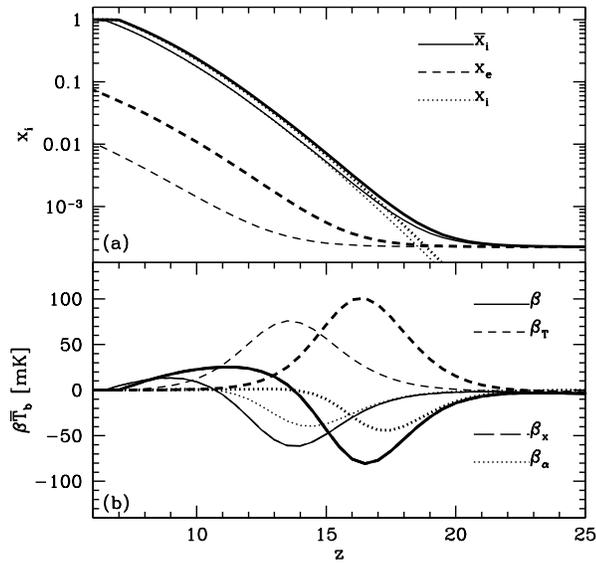


Figure 5.2: Ionization histories for model A (thick curves) and B (thin curves). (a): x_i (dotted curve), x_e (dashed curve), and the volume averaged ionization fraction $\bar{x}_i = x_i + (1 - x_i)x_e$ (solid curve). (b): The quantities $\beta_i \bar{T}_b$. We plot β (solid curve), β_x (long dashed curve, overlapping with β), β_T (short dashed curve), and β_α (dotted curve).

discussion of its effects until §5.6.3.

The global histories produced by these models are shown in Figure 5.1 and Figure 5.2. In Figure 5.1, we see the evolution of \bar{T}_K , \bar{T}_S , and T_γ . Note that, while both models produce the same qualitative behaviour, the reduced star formation rate in Model B delays the onset of heating from $z \approx 18$ to $z \approx 15$. We also see that the heating transition, where $\bar{T}_K = T_\gamma$, occurs at $z_h \approx 14$ in Model A and $z_h \approx 11$ in Model B. We have assumed that the X-ray luminosity per unit star formation is the same for both populations, so this is an effect of f_\star rather than the initial mass function. In addition, Pop. III stars produce fewer Ly α photons than Pop. II stars further slowing the onset of Lyman coupling.

Figure 5.1b shows the distinctive \bar{T}_b signature of absorption at $z > z_h$ followed by emission at $z < z_h$ in both models. The signal is significantly larger and more extended in Model A (See Furlanetto, 2006, for more detailed discussion of such histories). The ionization history is outlined in Figure 5.2a and shows that x_i evolves similarly in both models, as they have similar values for ζ . The electron fraction in the IGM x_e is depressed in model B, where there is a smaller X-ray background. Note that x_e remains much smaller than x_i once ionization begins. Both ionization histories produce an optical depth to the

surface of last scattering $\tau_{ri} \approx 0.07$, consistent with the WMAP third year observations of $\tau_{ri} = 0.09 \pm 0.03$ (Spergel et al., 2006), although slightly on the low side. Our model for temperature fluctuations will be geared towards making predictions for the largely neutral IGM outside of the ionized HII regions surrounding clusters of UV sources. Consequently, from Figure 5.2a, we expect our model to be valid for $z \gtrsim 12$, where $x_e \lesssim 0.1$ and the filling fraction of the HII regions is small.

Figure 5.2b shows $\beta_i \bar{T}_b$, which is a measure of the sensitivity of the 21 cm signal to fluctuations in each fundamental quantity. If the 21 cm signal were dominated by component i and if the fluctuation had unit amplitude $\delta_i \approx 1$, then $\beta_i \bar{T}_b$ gives the amplitude of the 21 cm signal. Note that the curves for δ and δ_x are almost indistinguishable and track \bar{T}_b . In contrast, the curves for δ_α and δ_T show clear peaks – representing windows where an existing signal might be seen. We may identify z_h as the point where $\bar{T}_b = 0$ and all curves except that for β_T go to zero. At this point, the only fluctuations in T_b arise from fluctuations in T_K . In practice, this “null” is more mathematical than physical, as inhomogeneities will blur the situation. The redshift window for observing the 21 cm signal is clearly much narrower in model B, indicating that it will be much more confused than in model A.

5.4 Formalism for temperature and ionization fluctuations

Having specified our global history, we now turn to calculating the fractional fluctuations δ_α , δ_T , and δ_x . Note that we will primarily be interested in the signal from the bulk of the IGM, working at redshifts where $x_i \lesssim 0.1$, so that we will ignore the fluctuations induced by HII regions. We begin by forming equations for the evolution of δ_T and δ_e (the fractional fluctuation in x_e) by perturbing equations (5.9) and (5.11) (see also Barkana & Loeb, 2005c; Naoz & Barkana, 2005). This gives

$$\frac{d\delta_T}{dt} - \frac{2}{3} \frac{d\delta}{dt} = \sum_i \frac{2\bar{\Lambda}_{\text{heat},i}}{3k_B \bar{T}_K} [\delta_{\Lambda_{\text{heat},i}} - \delta_T], \quad (5.27)$$

$$\frac{d\delta_e}{dt} = \frac{(1 - \bar{x}_e)}{\bar{x}_e} \bar{\Lambda}_e [\delta_{\Lambda_e} - \delta_e] - \alpha_A C \bar{x}_e \bar{n}_H [\delta_e + \delta], \quad (5.28)$$

where an overbar denotes the mean value of that quantity, and $\Lambda = \epsilon/n$ is the ionization or heating rate per baryon. We also need the fluctuation in the neutral fraction,

$\delta_x = -x_e/(1 - x_e)\delta_e$, and in the Ly α coupling coefficient $\delta_\alpha = \delta_{J_\alpha}$, neglecting the mild temperature dependence of S_α (Furlanetto & Pritchard, 2006).

To obtain a closed set of equations, we must calculate the fluctuations in the heating and ionizing rates. Perturbing equation (5.14) we find that the contribution of Compton scattering to the right-hand side of equation (5.27) becomes (Naoz & Barkana, 2005)

$$\frac{2\bar{\Lambda}_{\text{heat,C}}}{3k_B\bar{T}_K}[\delta\Lambda_{\text{heat,C}} - \delta_T] = \frac{\bar{x}_e}{1 + f_{\text{He}} + \bar{x}_e} \frac{a^{-4}}{t_\gamma} \left[4 \left(\frac{\bar{T}_\gamma}{\bar{T}_K} - 1 \right) \delta_{T_\gamma} + \frac{\bar{T}_\gamma}{\bar{T}_K} (\delta_{T_\gamma} - \delta_T) \right], \quad (5.29)$$

where δ_{T_γ} is the fractional fluctuation in the CMB temperature, and we have ignored the effect of ionization variations in the neutral fraction outside of the ionized bubbles, which are small. Before recombination, tight coupling sets $T_K = T_\gamma$ and $\delta_T = \delta_{T_\gamma}$. This coupling leaves a scale dependent imprint in the temperature fluctuations, which slowly decreases in time. We will ignore this effect, as it is small ($\sim 10\%$) below $z = 20$ and once X-ray heating becomes effective any memory of these early temperature fluctuations is erased. At low z , the amplitude of δ_{T_γ} becomes negligible, and equation (5.29) simplifies.

Our main challenge then is to calculate the fluctuations in the X-ray heating. We shall achieve this by paralleling the approach of Barkana & Loeb (2005b) to calculating fluctuations in the Ly α flux from a population of stellar sources. We first outline their results (see also Pritchard & Furlanetto, 2006). Density perturbations at redshift z' source fluctuations in J_α seen by a gas element at redshift z via three effects. First, the number of galaxies traces, but is biased with respect to, the underlying density field. As a result, an overdense region will contain a factor $[1 + b(z')\delta]$ more sources, where $b(z')$ is the (mass-averaged) bias, and will emit more strongly. Next, photon trajectories near an overdense region are modified by gravitational lensing, increasing the effective area by a factor $(1 + 2\delta/3)$. Finally, peculiar velocities associated with gas flowing into overdense regions establish an anisotropic redshift distortion, which modifies the width of the region contributing to a given observed frequency. Given these three effects, we can write $\delta_\alpha = \delta_{J_\alpha} = W_\alpha(k)\delta$, where we compute the window function $W_{\alpha,\star}(k)$ for a gas element at z by adding the coupling due to Ly α flux from each of the Ly n resonances and integrating over radial shells (Barkana & Loeb, 2005b)

$$W_{\alpha,\star}(k) = \frac{1}{J_{\alpha,\star}} \sum_{n=2}^{n_{\text{max}}} \int_z^{z_{\text{max}}(n)} dz' \frac{dJ_\alpha^{(n)}}{dz'} \frac{D(z')}{D(z)} \left\{ [1 + b(z')] j_0(kr) - \frac{2}{3} j_2(kr) \right\}, \quad (5.30)$$

where $D(z)$ is the linear growth function, $r = r(z, z')$ is the distance to the source, and the $j_l(x)$ are spherical Bessel functions of order l . The first term in brackets accounts for galaxy bias while the second describes velocity effects. The ratio $D(z')/D(z)$ accounts for the growth of perturbations between z' and z . Each resonance contributes a differential comoving Ly α flux $dJ_\alpha^{(n)}/dz'$, calculated from equation (5.24).

We plot $W_{\alpha,\star}(k)$ in Figure 5.3. On large scales, $W_{\alpha,\star}(k)$ approaches the average bias of sources, while on small scales it dies away rapidly, encoding the property that the Ly α flux becomes more uniform. In addition to the fluctuations in $J_{\alpha,\star}$, there will be fluctuations in $J_{\alpha,X}$. We calculate these below, but note in passing that the effective value of W_α is the weighted average $W_\alpha = \sum_i W_{\alpha,i}(J_{\alpha,i}/J_\alpha)$ of the contribution from stars and X-rays.

We now extend the formalism of Barkana & Loeb (2005b) in an obvious way to calculate fluctuations in the X-ray heating rate. First, note that for X-rays $\delta_{\Lambda_{\text{ion}}} = \delta_{\Lambda_{\text{heat}}} = \delta_{\Lambda_\alpha} = \delta_{\Lambda_X}$, as the rate of heating, ionization, and production of Ly α photons differ only by constant multiplicative factors (provided that we may neglect fluctuations in x_e , which are small). In each case, fluctuations arise from variation in the X-ray flux. We then write $\delta_{\Lambda_X} = W_X(k)\delta$ and obtain

$$W_X(k) = \frac{1}{\bar{\Lambda}_X} \int_{E_{\text{th}}}^{\infty} dE \int_z^{z_\star} dz' \frac{d\Lambda_X(E)}{dz'} \frac{D(z')}{D(z)} \left\{ [1 + b(z')]j_0(kr) - \frac{2}{3}j_2(kr) \right\}, \quad (5.31)$$

where the contribution to the energy deposition rate by X-rays of energy E emitted with energy E' from between redshifts z' and $z' + dz'$ is given by

$$\frac{d\Lambda_X(E)}{dz'} = \frac{4\pi}{h} \sigma_\nu(E) \frac{dJ_X(E, z)}{dz'} (E - E_{\text{th}}), \quad (5.32)$$

and $\bar{\Lambda}_X$ is obtained by performing the energy and redshift integrals. Note that rather than having a sum over discrete levels, as in the Ly α case, we must integrate over the X-ray energies. The differential X-ray number flux is found from equation (5.17).

The window function $W_X(k)$ gives us a “mask” to relate fluctuations to the density field; its scale dependence means that it is more than a simple bias. The typical sphere of influence of the sources extends to several Mpc. On scales smaller than this, the shape of $W_X(k)$ will be determined by the details of the X-ray source spectrum and the heating cross-section. On larger scales, the details of the heated regions remain unresolved so that

$W_X(k)$ will trace the density fluctuations.

A further word of explanation about this calculation is worthwhile. An X-ray is emitted with energy E' at a redshift z' and redshifts to an energy E at redshift z , where it is absorbed. To calculate W_X we perform two integrals in order to capture the contribution of all X-rays produced by sources at redshifts $z' > z$. The integral over z' counts X-rays emitted at all redshifts $z' > z$ which redshift to an energy E at z ; the integral over E then accounts for all the X-rays of different energies arriving at the gas element. Together these integrals account for the full X-ray emission history and source distribution. Many of these X-rays have travelled considerable distances before being absorbed. The effect of the intervening gas is accounted for by the optical depth term in J_X . Soft X-rays have a short mean free path and so are absorbed close to the source; hard X-rays will travel further, redshifting as they go, before being absorbed. Correctly accounting for this redshifting when calculating the optical depth is vital as the absorption cross-section shows strong frequency dependence. In our model, heating is dominated by soft X-rays, from nearby sources, although the contribution of harder X-rays from more distant sources cannot be neglected.

We compare the form of $W_X(k)$ and the stellar component of $W_\alpha(k)$ in Figure 5.3. Including the X-ray contribution in $W_\alpha(k)$ drives that curve towards the $W_X(k)$ curve. Note that W_X shows significantly more power on smaller scales than W_α , reflecting the greater non-uniformity in the X-ray heating; most heating comes from soft X-rays, which have mean free paths much smaller than the effective horizon of Ly α photons. Also, while W_α shows a subtle break in slope at $k \approx 3 \text{ Mpc}^{-1}$, W_X shows no obvious features indicative of preferred scales. Both W_X and W_α trace the bias on very large scales.

Returning now to the calculation of temperature fluctuations, to obtain solutions for equations (5.27) and (5.28), we let $\delta_T = g_T(k, z)\delta$, $\delta_e = g_e(k, z)\delta$, $\delta_\alpha = W_\alpha(k, z)\delta$, and $\delta_{\Lambda_X} = W_X(k, z)\delta$, following the approach of Bharadwaj & Ali (2004). Unlike Bharadwaj & Ali (2004), we do not assume these quantities to be independent of scale, and so we must solve the resulting equations for each value of k . Note that we do not include the scale dependence induced by coupling to the CMB (Naoz & Barkana, 2005). In the matter-dominated limit, we have $\delta \propto (1+z)^{-1}$ and so obtain

$$\frac{dg_T}{dz} = \left(\frac{g_T - 2/3}{1+z} \right) - Q_X(z)[W_X(k) - g_T] - Q_C(z)g_T, \quad (5.33)$$

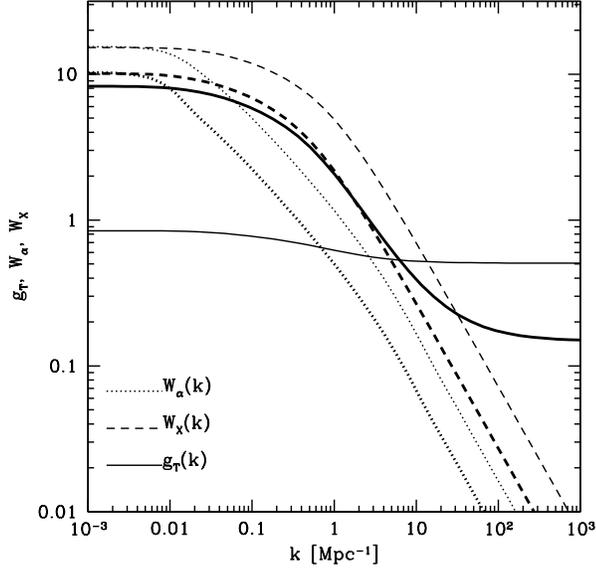


Figure 5.3: $W_{\alpha,*}(k)$ (dotted curves), $W_X(k)$ (dashed curves), and $g_T(k)$ (solid curves) at $z = 20$ (thin curves) and $z = 15$ (thick curves) for Model A.

$$\frac{dg_e}{dz} = \left(\frac{g_e}{1+z} \right) - Q_I(z)[W_X(k) - g_e] + Q_R(z)[1 + g_e], \quad (5.34)$$

where we define

$$Q_I(z) \equiv \frac{(1 - \bar{x}_e)}{\bar{x}_e} \frac{\bar{\Lambda}_{\text{ion},X}}{(1+z)H(z)}, \quad (5.35)$$

$$Q_R(z) \equiv \frac{\alpha_A C \bar{x}_e \bar{n}_H}{(1+z)H(z)}, \quad (5.36)$$

$$Q_C(z) \equiv \frac{\bar{x}_e}{1 + f_{\text{He}} + \bar{x}_e} \frac{(1+z)^3 T_\gamma}{t_\gamma H(z) \bar{T}_K}, \quad (5.37)$$

and

$$Q_X(z) \equiv \frac{2\bar{\Lambda}_{\text{heat},X}}{3k_B \bar{T}_K (1+z)H(z)}. \quad (5.38)$$

These are defined so that Q_R and Q_I give the fractional change in x_e per Hubble time as a result of recombination and ionization respectively. Similarly, Q_C and Q_X give the fractional change in \bar{T}_K per Hubble time as a result of Compton and X-ray heating. Immediately after recombination Q_C is large, but it becomes negligible once Compton heating becomes ineffective at $z \sim 150$. The Q_R term becomes important only towards the end of reionization, when recombinations in clumpy regions slows the expansion of HII regions. Only the Q_X and Q_I terms are relevant immediately after sources switch on. We must integrate these

equations to calculate the temperature and ionization fluctuations at a given redshift and for a given value of k .

These equations illuminate the effect of heating. First, consider g_T , which we can easily relate to the adiabatic index of the gas γ_a by $g_T = \gamma_a - 1$, giving it a simple physical interpretation. Adiabatic expansion and cooling tends to drive $g_T \rightarrow 2/3$ (corresponding to $\gamma_a = 5/3$, appropriate for a monoatomic ideal gas), but when Compton heating is effective at high z , it deposits an equal amount of heat per particle, driving the gas towards isothermality ($g_T \rightarrow 0$). At low z , where X-ray heating of the gas becomes significant, the temperature fluctuations are dominated by spatial variation in the heating rate ($g_T \rightarrow W_X$). This embodies the higher temperatures closer to clustered sources of X-ray emission. If the heating rate is uniform $W_X(k) \approx 0$, then the spatially constant input of energy drives the gas towards isothermality ($g_T \rightarrow 0$).

The behaviour of g_e is similarly straightforward to interpret. At high redshift, when the IGM is dense and largely neutral, the ionization fraction is dominated by the recombination rate, pushing $g_x \rightarrow -1$, because denser regions recombine more quickly. As the density decreases and recombination becomes ineffective, the first term of equation (5.34) slowly pushes $g_x \rightarrow 0$. Again, once ionization becomes important, the ionization fraction is pushed towards tracking spatial variation in the ionization rate ($g_x \rightarrow W_X$). Note that, because the ionization fraction in the bulk remains less than a few percent, fluctuations in the neutral fraction remain negligibly small at all times.

The scale dependence of g_T is illustrated in Figure 5.3. g_T tries to track the heating fluctuations $W_X(k)$ (as in the $z = 15$ curve), but two factors prevent this. First, until heating is significant, the effect of adiabatic expansion tends to smooth out variations in g_T . Second, g_T responds to the integrated history of the heating fluctuations, so that it tends to lag W_X somewhat. When the bulk of star formation has occurred recently, as when the star formation rate is increasing with time, then there is little lag between g_T and W_X . In contrast, when the star formation rate has reached a plateau or is decreasing, the bulk of the X-ray flux originates from noticeably higher z , and so g_T tends to track the value of W_X at this higher redshift. On small scales, the heating fluctuations are negligible and g_T returns to the value of the (scale-independent) uniform heating case.

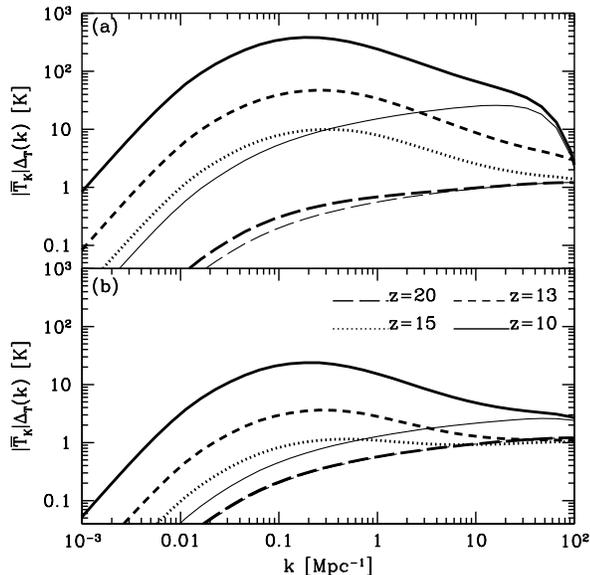


Figure 5.4: Fluctuations in T_K . In each panel we plot $\bar{T}_K \Delta_T(k)$ for the case of inhomogeneous X-ray heating (thick curves) at $z = 20$ (long dashed curve), $z = 15$ (dotted curve), $z = 13$ (short dashed curve), and $z = 10$ (solid curve). For comparison, we plot the case of uniform heating at $z = 10$ (thin solid curve) and $z = 20$ (thin long dashed curve). (a): Model A. (b): Model B.

5.5 Temperature fluctuations

Before calculating the 21 cm signal, let us first examine the gas temperature fluctuations themselves. Figure 5.4 shows the power spectrum of temperature fluctuations $P_T(k)$ for models A and B respectively. We see that in both cases the fluctuations are small until $z < 20$. At lower redshifts and on larger scales ($k \approx 0.1 Mpc^{-1}$), the heating fluctuations source a significant (factor of ≈ 50) enhancement over the uniform heating case. This is to be expected. Uniform heating of the gas tends to erase temperature fluctuations, while inhomogeneous heating causes them to grow. Thus we observe a huge increase in power. The fluctuation amplitude in Model B is generally smaller than in Model A as a consequence of the reduced heating from the decreased SFR in Model B. In both cases, the temperature fluctuations remain small, $\delta_T < 1$ (compare with Figure 5.1), justifying our linear approximations.

Figure 5.5 illustrates the redshift evolution of the temperature fluctuations. We choose to follow a single wavenumber $k = 0.1 Mpc^{-1}$, which is both within those scales accessible to

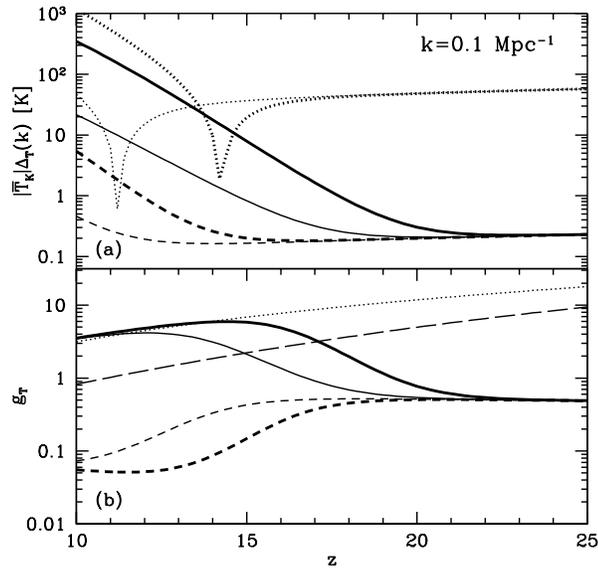


Figure 5.5: Evolution of the fluctuations in T_K with redshift for Model A (thick curves) and B (thin curves). (a): We plot the amplitude of fluctuations in T_K given by $\bar{T}_K \Delta_T(k)$ at $k = 0.1 \text{ Mpc}^{-1}$ in the case of uniform heating (dashed curves) and when fluctuations in the heating rate are considered (solid curves). For comparison, we plot $|\bar{T}_k - T_\gamma|$ (dotted curves). Only in a small region of width $\Delta z \approx 1$ around $\bar{T}_K = T_\gamma$ do the fluctuations exceed this threshold. (b): Evolution of g_T . We plot g_T at $k = 0.1 \text{ Mpc}^{-1}$ for the uniform (short dashed curves) and fluctuating cases (solid curves). We also plot W_X (dotted curve) and, for comparison, $W_{\alpha,*}(k)$ (long dashed curve). Notice how g_T rises to track W_X once heating becomes effective.

future experiments and demonstrative of the effect. If the gas is heated uniformly (dashed curves), then g_T rapidly becomes negligible once heating becomes effective. By depositing the same amount of energy per particle the gas is driven towards isothermality. When heating fluctuations are taken into account g_T may grow or decrease depending on scale. We observe that, for the scale chosen here, the amplitude of the temperature fluctuations grows steadily with time, but g_T decreases. This is a consequence of the sources becoming less biased with time so that $W_X(z)$ decreases with z . On very small scales, where $W_X(k)$ is negligible, g_T will trace the uniform heating curve.

Recall that whether we observe the 21 cm line in emission or absorption depends on the sign of $T_S - T_\gamma$. Assuming that $T_S \approx T_K$, when $T_K < T_\gamma$, hotter regions have a spin temperature closer to the CMB temperature and so appear more faintly in absorption. As heating continues, it is these regions that are first seen in emission, when their temperature exceeds T_γ . Once $T_K > T_\gamma$, these hotter regions produce the largest emission signal.

We see from Figure 5.5 that for a short window around z_h (where $\bar{T}_K = T_\gamma$) temperature fluctuations may raise T_K above T_γ in these hot regions, even when \bar{T}_K is less than T_γ . We interpret this to mean that within this window the 21 cm signal will be a confusing mix of emission, from hotter regions, and absorption, from cooler regions. In the case of uniform heating this window is very narrow, but when fluctuations are included it extends to a significant ($\Delta z \approx 1$) width. This indicates that the transition from absorption to emission will not be abrupt, but extended.

5.6 21 cm power spectrum

5.6.1 Redshift evolution

Finally, we write the full 21 cm power spectrum as

$$P_{T_b}(k, \mu) = \bar{T}_b^2 (\beta' + \mu^2)^2 P_{\delta\delta}(k), \quad (5.39)$$

where

$$\beta' = \beta - \beta_x \bar{x}_e g_e / (1 + \bar{x}_e) + \beta_T g_T + \beta_\alpha W_\alpha. \quad (5.40)$$

Within our model we may neglect the term corresponding to the neutral fraction, as the free-electron fraction in the IGM remains small at all times. We now consider how the 21

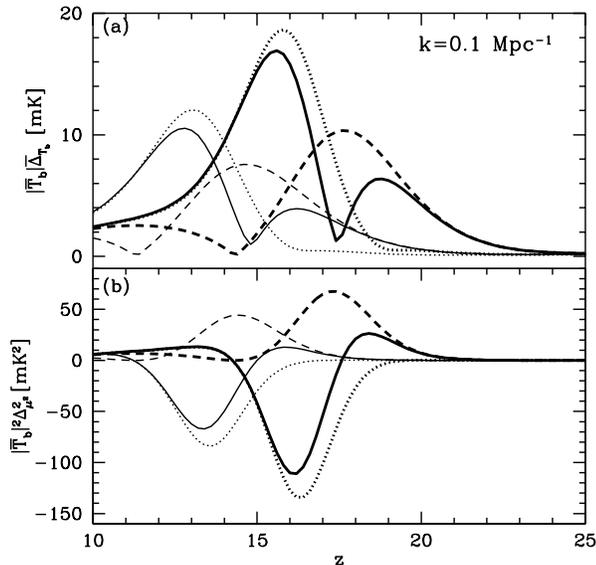


Figure 5.6: Evolution of brightness temperature fluctuations for Model A (thick curves) and B (thin curves). (a): We plot $|\bar{T}_b| \bar{\Delta}_{T_b}(k)$ at $k = 0.1 \text{ Mpc}^{-1}$ including the effects of heating fluctuations (dotted curves), Ly α fluctuations (dashed curves), and both heating and Ly α fluctuations (solid curves). (b): We plot $|\bar{T}_b|^2 \Delta_{\mu^2}^2(k)$ with the same line conventions.

cm power spectrum evolves with redshift.

Figure 5.6 shows the evolution of the brightness temperature fluctuations at a single scale $k = 0.1 \text{ Mpc}^{-1}$ with redshift. First, note that in the bottom panel $\Delta_{\mu^2}^2$ changes sign when we include temperature fluctuations (note that $\Delta_{\mu^2}^2$ is not an auto-correlation and so is free to have a negative sign). Physically, this occurs because when $T_K < T_\gamma$ there is an anti-correlation between T_b and T_K , i.e. increasing T_K decreases T_b . Observing $P_{\mu^2} < 0$ is a clear sign that $T_K < T_\gamma$. Mathematically, this can be seen because β_T is the only one of the fluctuation coefficients that can become negative. Of course, if $P_{\delta T}$ or other cross-correlations become negative we can also get $P_{\mu^2} < 0$, but this should not be the case for radiative heating or Ly α coupling, as we expect emitting sources to be most common in overdense regions. Only in the case of $P_{x\delta}$ might we expect a negative cross-correlation, as increasing the UV radiation is likely to decrease the neutral fraction. In the high redshift regime, before significant ionization has occurred, this term is negligible.

Adding the Ly α fluctuations, we see a clear double-peaked temporal structure in the evolution of $\bar{\Delta}_{T_b}$, which is dominated by Ly α fluctuations at high z and temperature fluc-

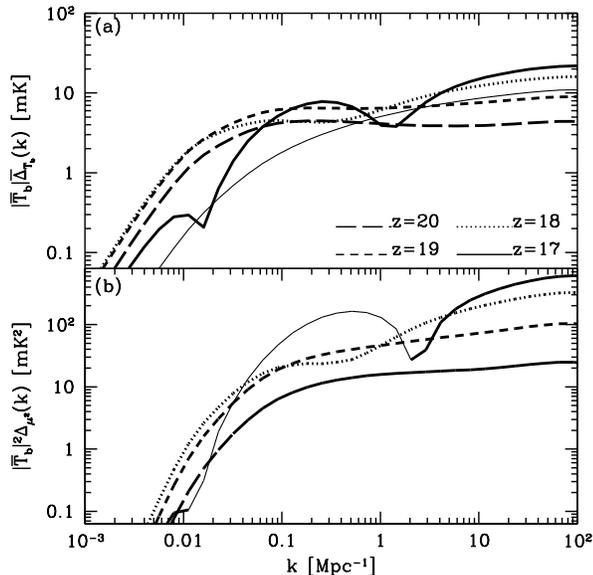


Figure 5.7: Full T_b power spectra for Model A. We plot the power spectra at $z = 20$ (long dashed curve), $z = 19$ (short dashed curve), $z = 18$ (dotted curve), and $z = 17$ (solid curve). (a): $|\bar{T}_b|\bar{\Delta}_{T_b}(k)$. We plot $|\bar{T}_b|\Delta_{\delta\delta}$ (thin solid curve) at $z = 19$ for comparison. (b): $|\bar{T}_b|^2\Delta_{\mu^2}^2(k)$. The sign of $\Delta_{\mu^2}^2(k)$ is indicated as positive (thick curves) or negative (thin curves).

tuations at lower z (were we to include the effects of ionization fluctuations, there would be a third peak at still lower redshift). We note that there is considerable overlap between the two signals, which will complicate extracting astrophysical information. The situation is similar in Model B, although here the relevant signal is compressed into a narrower redshift window. We note that the amplitude of fluctuations induced by the gas temperature is significantly larger than those from the Ly α signal and present at lower redshifts. Both of these features make the temperature fluctuation signal a plausible target for future observations.

To illustrate the scale dependence of this signal, we examine a series of redshift slices. We will make plots for model A. Although the same evolution applies for model B, the events are shifted to lower redshift $\Delta z \approx 3$ and the transitions are somewhat compressed in redshift. We begin by examining the high redshift regime, where Ly α fluctuations dominate the 21 cm signal, but temperature fluctuations become important as we move to lower redshift.

Figure 5.7 shows redshift slices from $z = 17 - 20$. We can see from Figure 5.6 that Ly α fluctuations dominate the signal for $z \gtrsim 18$. The $z = 20$ and $z = 19$ curves show the expected excess of power on large scales for Ly α fluctuations from the first sources (see

Barkana & Loeb, 2005b, for a full analysis of this signal). At $z = 18$, we begin to see the effects of the temperature fluctuations through the dip in power between $k = 0.1$ and 1 Mpc^{-1} . This dip occurs because $\beta_T < 0$, contrasting with the other β_i , which are positive. Physically, in this regime $T_K < T_\gamma$ and regions that are hotter have a smaller brightness temperature. In our model, denser regions are more strongly coupled, which increases T_b , but are also hotter, which tends to decrease T_b . These two effects compete with one another and produce the dip.

At $z = 17$, temperature fluctuations grow large enough to drive β' negative over a range of scales, where they outweigh the Ly α fluctuations. This leads to a sign change in $\Delta_{\mu^2}^2$, but also imprints a distinctive trough-peak-trough structure in $\bar{\Delta}_{T_b}$. Here Ly α fluctuations dominate on the largest scales, temperature fluctuations on intermediate scales, and density fluctuations on small scales. For this to occur, we require that $W_\alpha > g_T$ on large scales, which can only occur if W_α and g_T show different scale dependence. This always occurs at some redshift in our model, as both W_α and W_X tend towards the same value on large scales, but g_T lags behind (and so is smaller than) W_X on those scales.

From Figure 5.6, we see that T_K fluctuations dominate at $z \lesssim 17$ and that Ly α fluctuations become negligible for $z \lesssim 15$. In Figure 5.8, we plot redshift slices in the range $z = 13 - 16$. At $z = 16$ and $z = 15$, we see a sign change in Δ_{μ^2} , which is a distinctive signature of the temperature fluctuations when $T_K < T_\gamma$. This is seen in $\bar{\Delta}_{T_b}$ as a peak on large scales, followed by a trough at smaller scales. The position of the peak depends upon the shape of g_T and thus the X-ray source spectrum. We will consider this in more detail in the next section.

Notice that the heating transition occurs very close to $z = 14$, so that the 21 cm signal at this redshift would likely be seen in a mixture of absorption and emission. In addition, this curve is dominated by gas temperature fluctuations. We see this in Figure 5.8 where the contribution from density fluctuations at $z = 14$ (thin solid curve) is at least a factor of two smaller than $\bar{\Delta}_{T_b}$ on all scales. Recall from Figure 5.2 that when $\bar{T}_b \approx 0$ only the combination $|\bar{T}_b|\beta_T$ is significant.

The position of the sign change moves to smaller scales as the gas is heated and the temperature fluctuations become larger. Eventually, the IGM heats to $\bar{T}_K > T_\gamma$, hotter regions have a higher brightness temperature than average and $\beta_T > 0$. Once this occurs the trough disappears entirely and the peak on large scales is no longer quite so distinctive

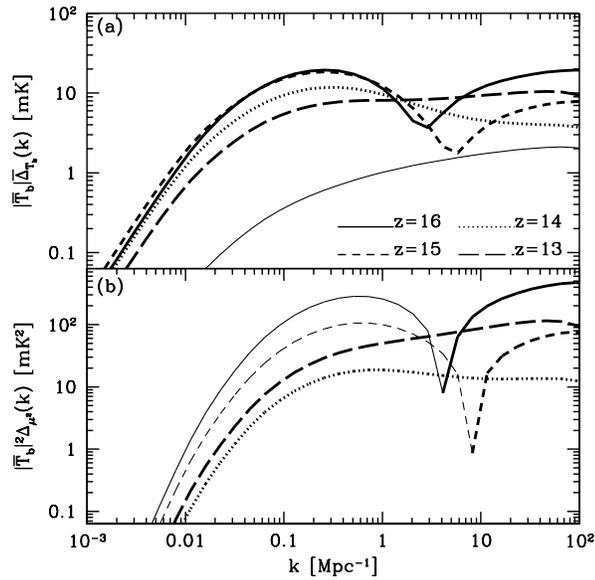


Figure 5.8: Full T_b power spectra for Model A. We plot the power spectra at $z = 16$ (solid curve), $z = 15$ (short dashed curve), $z = 14$ (dotted curve), and $z = 13$ (long dashed curve). Note that the $z = 13$ curve would be seen in emission, and the $z = 14$ curve in a mixture of emission and absorption. The other curves would be seen in absorption against the CMB. (a): $|\bar{T}_b| \bar{\Delta}_{T_b}(k)$. We plot $|\bar{T}_b| \Delta_{\delta\delta}$ (thin solid curve) at $z = 14$ for comparison. (b): $|\bar{T}_b|^2 \Delta_{\mu^2}^2(k)$. The sign of $\Delta_{\mu^2}^2(k)$ is indicated as positive (thick curves) or negative (thin curves).

(see $z = 14$ curve). The continued IGM heating drives $\beta_T \rightarrow 0$ and diminishes the effect of the temperature fluctuations. By $z = 13$, there is no longer a clear peak in either $\Delta_{\mu^2}^2$ or $\bar{\Delta}_{T_b}$, although there is still considerable excess power on large scales. By $z = 10$, $T_K \gg T_\gamma$ and temperature fluctuations no longer impact the 21 cm signal significantly.

Once the ionization fraction becomes large ($x_i \gtrsim 0.1$), the 21 cm signal becomes dominated by the imprint of HII regions (Zaldarriaga et al., 2004; Furlanetto et al., 2004). This eventually produces a distinct knee in the 21 cm power spectrum resulting from the characteristic size of the bubbles. We note that our models have $x_i \lesssim 0.1$ at $z \gtrsim 12$, so that we do not expect ionization fluctuations to significantly affect the results we have outlined for Model A. In the case of Model B, temperature fluctuations remain significant to lower redshift where they may interfere with attempts to measure the power spectrum of ionization fluctuations. The reverse is also true.

5.6.2 Spectral dependence

We next imagine using the temperature fluctuations to constrain the X-ray source spectra. This should affect the temperature fluctuations on intermediate scales, where heating fluctuations dominate. Increasing the hardness of the spectrum increases the fraction of more energetic photons, which have longer mean free paths. This should further smooth the temperature fluctuations and suppress power on small scales.

Figure 5.9 shows the power spectra at $z = 15$ (chosen to maximise the distinctive features of the temperature fluctuations) for source spectra $\alpha_S = 1.5$ (mini-quasars), $\alpha_S = 1.0$ (SNR), and $\alpha_S = 0.5$. We see that the spectra alter the most on scales $k \approx 0.1 - 10 \text{ Mpc}^{-1}$. The two main signatures are the change in amplitude and shift in the position of the trough. Both of these occur because increasing the slope of the spectrum, with fixed total luminosity, increases the number of soft X-rays and so increases the heating in smaller scales. The trough (or sign change in P_{μ^2}) shifts by $\Delta k \sim 2 \text{ Mpc}^{-1}$ for $\Delta\alpha = 0.5$, an effect that might be observable were it not located on small scales $k \approx 5 \text{ Mpc}^{-1}$. The amplitude change at the peak is more observable but is also degenerate with modifications in the thermal history, making this a very challenging measurement to perform in practice.

Referring back to our discussion of the time evolution of the signal, we see that this sort of variation is similar to the effect of changing the thermal history. However, the exact shape of the spectrum is determined by the form of g_T , and hence W_X . These do encode

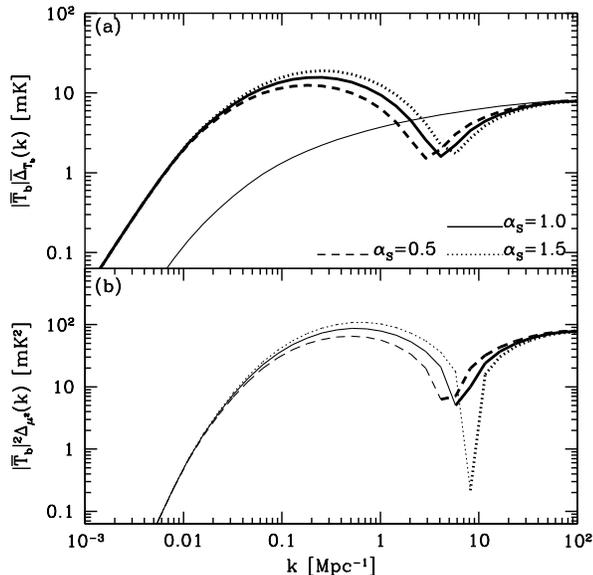


Figure 5.9: Effect of X-ray spectra on 21 cm power spectra. We show results at $z = 15$ for model A and take $\alpha = 1.5$ (dotted curve), $\alpha = 1.0$ (solid curve), and $\alpha = 0.5$ (dashed curve). (a): $|\bar{T}_b| \bar{\Delta}_{T_b}(k)$. We illustrate the uniform heating case by the thin solid curve. (b): $|\bar{T}_b|^2 \bar{\Delta}_{\mu^2}^2(k)$.

distinct information about the source spectrum. Consequently, precision measurements of the 21 cm power spectrum at high z could constrain the X-ray source spectrum.

We can also seek to constrain the X-ray spectrum by looking at the regime where fluctuations in the Ly α flux dominate the 21 cm signal. The inclusion of Ly α photons generated by X-ray excitation of HI (in addition to those redshifting into the Lyman resonances) modifies the shape of the power spectrum significantly. This is easy to see by referring back to Figure 5.3. There we plotted $W_{\alpha,\star}(k)$, for the case of stellar emission, and $W_X(k)$, which determines the fluctuations in the X-ray flux. If we allow both stars and Ly α photons produced from X-rays to contribute to the Ly α flux, then the resulting spectrum of fluctuations is determined by a weighted combination of these $W_\alpha(k)$ and $W_X(k)$. In our model, as in that of Chen & Miralda-Escude (2006), the Ly α flux is dominated on small scales by the X-ray contribution and on large scales by the stellar contribution. Thus the resulting weighting function most closely resembles $W_X(k)$ with significant power on small scales.

Figure 5.10 shows the effect on the power spectrum at $z = 20$, when temperature

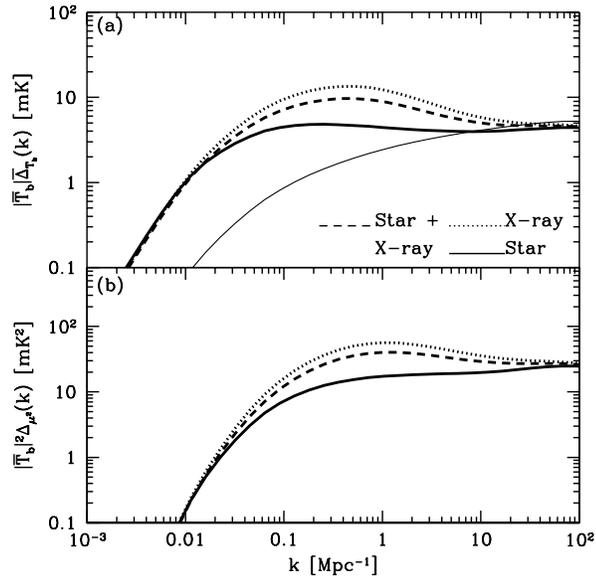


Figure 5.10: (a): $|\bar{T}_b|\bar{\Delta}_{T_b}(k)$. We consider the following sources of Ly α emission: stellar only (solid curve), X-ray excitation only (dotted curve), stellar+X-ray excitation (dashed curve). All curves are calculated at $z = 20$ and have been normalised to the stellar only case, to compensate for different mean values of x_α . We assume X-ray emission from starburst galaxies. Also plotted is $|\bar{T}_b|\Delta_{\delta\delta}$ (thin solid curve). (b): $|\bar{T}_b|^2\Delta_{\mu^2}^2(k)$. Same line conventions as in (a).

fluctuations are negligible, of including the different contributions to the Ly α flux. On intermediate scales ($k \approx 1 \text{ Mpc}^{-1}$) there is clearly significantly more power when X-ray excitation dominates Ly α production compared to stellar production. As noted in Chuzhoy et al. (2006), this provides a means for distinguishing between the major source of Ly α photons during the time of the first sources. We note that the shape of the spectrum is somewhat sensitive to the spectral index of the X-ray sources, with the variation being similar to between the stellar + X-ray and X-ray only curves. Thus isolating the 21 cm fluctuations from the Ly α flux variations could also constrain the X-ray spectrum of the first sources.

5.6.3 Effects of X-ray background

We now explore the effect of modifying the X-ray luminosity of our sources. We have so far taken $f_X = 1$ in our analysis, but constraints on the high redshift X-ray background are weak, giving us significant freedom to vary f_X , which parametrizes the source luminosity. As an example, for our model A, values of $f_X \lesssim 10^3$ are easily possible without X-ray or collisional ionization of the IGM violating WMAP3 constraints on τ at the 2-sigma level. In Figure 5.11, we show the time evolution of the 21 cm fluctuations for model A, taking $f_X = 0.1, 1, \text{ and } 10$. This serves to illustrate the effect of late or early X-ray heating and illustrates the range of uncertainty in making predictions.

Earlier heating (dashed curve) causes the temperature fluctuations to become important at higher redshift, cutting into the region of Ly α fluctuation. This will make the 21 cm signal more complicated as temperature and Ly α fluctuations contribute over a similar range of redshifts. However, early heating also means that temperature fluctuations become unimportant for the 21 cm signal at late times improving the prospects for extracting cosmology from the 21 cm signal (McQuinn et al., 2005; Santos et al., 2005). In contrast, late heating (dotted curve) allows a clearer separation between temperature and Ly α fluctuations, but means temperature fluctuations are likely to be important during the beginning of reionization. This will complicate the extraction of information about HII regions as reionization gets underway.

Clearly there is considerable uncertainty as to the behaviour of the 21 cm signal at high redshifts due to our poor understanding of the source populations. Viewed another way, measurement of the evolution of the 21 cm signal could provide useful constraints on the

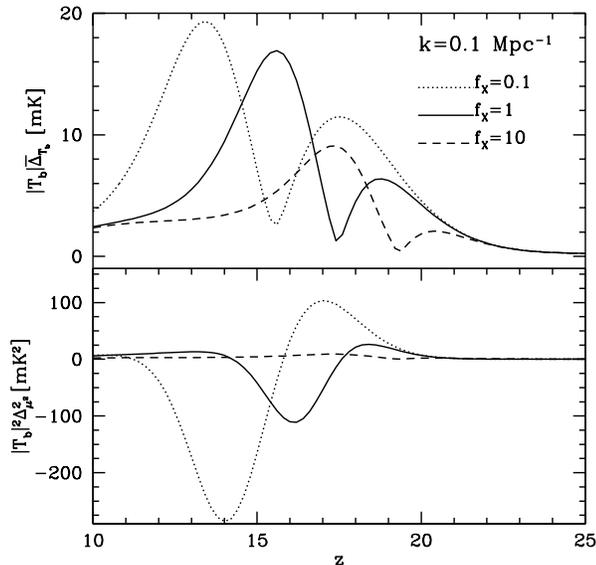


Figure 5.11: (a): Redshift evolution of $|\bar{T}_b|\bar{\Delta}_{T_b}(k)$ at $k = 0.1 \text{ Mpc}^{-1}$ for model A, but with $f_X = 0.1$ (dotted curve), 1.0 (solid curve), and 10 (dashed curve) (b): Redshift evolution of $|\bar{T}_b|\Delta_{\mu^2}(k)$. Same line conventions as in (a).

X-ray background at high redshift. This is important as efforts to observe the diffuse X-ray background are complicated by technical issues of calibration. We also note that for weaker X-ray heating other sources of heating, especially shock heating, may become important.

Finally we remind the reader that our model is applicable in the IGM outside of ionized HII regions. If heating occurs late, so that temperature fluctuations are important as HII regions become large, then it will be important to extend this model if accurate predictions of the 21 cm signal during reionization are to be made. It will also be important to include these temperature fluctuations in simulated predictions of the 21 cm signal.

5.7 Observational Prospects

We now turn to the important question of observing the features outlined above. The first generation of 21 cm experiments (PAST, LOFAR, MWA) will be optimised to look for the signature of HII regions at redshifts $z \lesssim 12$. Their sensitivity decreases rapidly at redshifts $z \gtrsim 10$ (McQuinn et al., 2005; Bowman et al., 2006) and so they are unlikely to be able to detect the effects of inhomogeneous heating. The proposed successor to these instruments, the SKA, is still under design, but its fiducial specifications should allow the $z > 12$ regime

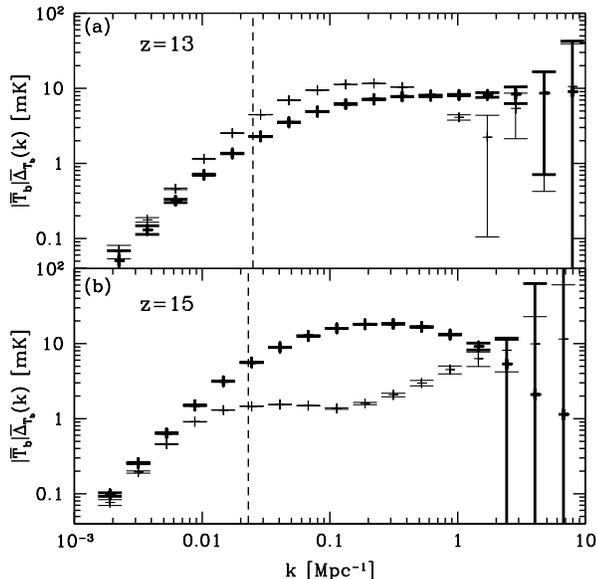


Figure 5.12: Predicted $1-\sigma$ errors on $|\bar{T}_b|\bar{\Delta}_{T_b}(k)$ for an SKA like instrument (see text for details). We compare spectra for model A (thick curves) and model B (thin curves). Modes with $k \lesssim k_{\text{foreground}}$ (shown by a vertical dashed line) will probably be lost during foreground cleaning. (a): $z = 13$. (b): $z = 15$.

to be probed. In this section, we will consider using an SKA type experiment to observe 21 cm fluctuations at $z = 13$ and $z = 15$ and calculate the achievable precision.

Before this, we must make the necessary caveats concerning foregrounds. Foregrounds for 21 cm observations include terrestrial radio interference (RFI), galactic synchrotron emission, radio recombination lines, and many others (Oh & Mack, 2003; Di Matteo et al., 2004; Furlanetto et al., 2006, §9). Typical foregrounds produce system temperatures $T_{\text{sys}} \gtrsim 1000$ K, compared to a signal measured in mK. These foregrounds increase rapidly as we move to lower frequency, making their removal an even greater concern for high redshift observations than low ones. Although techniques for foreground removal are well grounded, their effectiveness has yet to be tested. In the analysis that follows, we assume that foreground removal can be effected by exploiting the smoothness of foregrounds in frequency space (Zaldarriaga et al., 2004; Santos et al., 2005; Morales & Hewitt, 2004; Morales, 2005; McQuinn et al., 2005; Wang & Hu, 2006).

Figure 5.12 shows predicted $1-\sigma$ error bars on $\bar{\Delta}_{T_b}(k)$ at $z = 13$ and $z = 15$ for models A and B. We assume an SKA-like instrument with a total effective area $A_{\text{tot}} = 1 \text{ km}^2$ distributed over 5000 antennae in a 5 km core, bandwidth $B = 12 \text{ MHz}$, minimum baseline

$D_{\min} = 10$ m, and an integration time $t_{\text{int}} = 1000$ hr. We set $T_{\text{sys}} = 1000$ K and 1400 K at these two redshifts and use bins of width $\Delta k = k/2$. We assume that foregrounds can be removed exactly, but that this also removes cosmological information on scales exceeding the bandwidth of the observations, so that modes with $k \leq k_{\text{foreground}} \approx 0.025 \text{ Mpc}^{-1}$ (indicated by vertical dashed lines) are lost (McQuinn et al., 2005).

With these caveats, observations could measure $\bar{\Delta}_{T_b}(k)$ accurately over the range $k \approx 0.025\text{--}2 \text{ Mpc}^{-1}$. The precision is more than adequate to distinguish between model A and B. Detecting the characteristic peak-trough signature of δ_T is difficult, as the trough typically occurs on small scales where the uncertainty is large. However, it should be possible to detect the peak and the beginning of the decline. We note that detection of the trough is necessary to unambiguously determine $\beta_T < 0$ and so show that $T_S < T_\gamma$. Without this it is not simple to distinguish between the two cases exemplified by the $z = 14$ curve, which has no trough, and the $z = 15$ curve, which does (Figure 5.8). No similar confusion occurs when the reduction in power caused by $\beta_T < 0$ is obvious, as in the $z = 15$ model B case.

From the point of view of constraining the spectra of X-ray sources, the precision is adequate for distinguishing between the different curves of Figure 5.10. Whether the effect of the spectrum can be separated out from different thermal histories is an open question that deserves future study.

Throughout this work, we have ignored the effect of the HII regions on the 21 cm power spectrum. While this is reasonable at high redshifts, this approximation will begin to break down as the filling fraction of ionized regions increases. The bubble model of Furlanetto et al. (2004) predicts that these bubbles remain at sub-Mpc sizes while $x_i \lesssim 0.1$. Consequently, we naively expect contamination of the signal by these bubbles to be confined to small scale modes with $k \gtrsim 1 \text{ Mpc}^{-1}$ that will be very difficult to detect. Exploring the detailed interaction between temperature and neutral fraction fluctuations is beyond the scope of this paper, but may be important for detailed predictions of the 21 cm signal at the beginning of reionization.

Santos & Cooray (2006) have considered the extraction of astrophysical and cosmological parameters from 21 cm observations in the period of the first sources. They assumed that gas temperature fluctuations showed no scale dependence $g_T(k, z) = g_T(z)$ and argued that extracting astrophysical information using an SKA like instrument would be difficult but is feasible. We expect the scale dependent temperature fluctuations that we have investigated

to both help and hinder parameter estimation. Figure 5.12 shows that it should be possible to resolve individual features imprinted in the power spectrum by temperature fluctuations. These features provide additional leverage in extracting astrophysical parameters. However, the shape of the power spectrum evolves rapidly in our model, making binning of different redshift data more difficult.

5.8 Conclusions

X-ray production by an early generation of stellar remnants is widely regarded as the most likely candidate for heating the IGM above the CMB temperature from its cool adiabatic level. This heating has often been treated as uniform, as the mean free path of hard X-rays in the early Universe is comparable to the Hubble scale. We have relaxed this assumption and, by expanding on the formalism of Barkana & Loeb (2005b), calculated the temperature fluctuations that arise from the inhomogeneous heating. The spectrum of fluctuations in T_K is significantly larger than that predicted from uniform heating, peaking on scales $k \approx 0.1 \text{ Mpc}^{-1}$. This allowed us to examine the redshift range about z_h , where $T_K = T_\gamma$, and show that there is a window of width $\Delta z \approx 1$ in which the IGM will contain pockets of gas both hotter and colder than the CMB. This has implications for the 21 cm signal, which will be seen in a mixture of absorption and emission within this window.

The best hope for observing the temperature evolution before reionization is through 21 cm observations of neutral hydrogen. Systematic effects arising from foregrounds are likely to prevent interferometers from measuring \bar{T}_b directly (Furlanetto et al., 2006), although several alternative methods for obtaining \bar{T}_b have been proposed (Barkana & Loeb, 2005a; Cooray, 2006). Thus careful analysis of brightness fluctuations will be required to extract astrophysical information. Fluctuations in T_K lead to fluctuations in T_b , which contain information about the thermal history and the nature of the heating sources. We have calculated the 21 cm power spectrum arising from inhomogeneous X-ray heating and shown that it has considerable structure. In the regime where gas temperature and Ly α flux fluctuations compete, we expect a trough-peak-trough structure in $\bar{\Delta}_{T_b}(k)$. Once T_K fluctuations dominate, but while $T_K < T_\gamma$, we see a peak-trough structure. As the gas heats, this structure is lost as the trough moves to unobservable small scales while the peak decreases and finally vanishes once $T_K \gg T_\gamma$. Extracting astrophysical information cleanly

will be challenging, but the information is there.

It is important to notice that the difference between uniform and inhomogeneous heating is large. Observations with the SKA should be able to distinguish these two cases and indicate whether X-ray heating is important. If it is possible to perform an angular separation of P_{T_b} , then observing $P_{\mu^2} < 0$ is a clear indicator that $T_K < T_\gamma$. Ideally, one would extract the quantity β_T , but this requires fitting of other parameters and so is a less direct (but more conclusive) observational feature.

Additionally, the spectra of the X-ray sources imprint information on the T_K fluctuations. This may be observed in the 21 cm power spectra, where it shifts the critical scale at which P_{μ^2} changes sign, or during the regime in which Ly α fluctuations dominate, where it modifies the shape of the power spectrum. The temperature fluctuations that we have calculated lead to a 21 cm signal that extends down to relatively low redshifts. This opens an opportunity for future 21 cm radio arrays to probe the thermal history prior to reionization. Including temperature fluctuations makes the 21 cm signal significantly more complex, adding information, but further raises the question of how best to separate out that information.

In this paper, we have ignored the contribution from Poisson fluctuations in the source distribution (Barkana & Loeb, 2005b). While calculating it requires only a straightforward extension of the Barkana & Loeb (2005b) formalism, performing the time integrals necessary to convert heating fluctuations into temperature fluctuations is non-trivial. We have estimated the amplitude of these Poisson temperature fluctuations and find them (in our models) to be subdominant at all redshifts. This is largely because there are many more sources at the lower redshifts where temperature fluctuations are important. In theory, high-precision 21 cm observations can separate these Poisson fluctuations from fluctuations correlated with the density field. The Poisson contribution could then be used to probe the distribution of sources, for example, by distinguishing between highly biased mini-quasars and less biased star-burst galaxies, producing the same global X-ray luminosity.

In our analysis we have taken $f_X = 1$, corresponding to normalising the X-ray luminosity per unit star formation to that observed in the local universe. In truth, this assumption is highly speculative and the value for f_X is extremely uncertain. We have investigated the effects of changing f_X and find that it alters the details of the thermal evolution significantly. Taking $f_X = 0.1$, for example, shifts the point where 21 cm brightness fluctuations change

from being dominated by Ly α fluctuations to gas temperature fluctuations from $z \approx 17$ to $z \approx 15$. Setting $f_X = 10$ increases the redshift of this transition to $z \approx 19$. For values of $f_X \lesssim 0.1$, we find a clear separation between 21 cm brightness fluctuations sourced by gas temperature and Ly α fluctuations. Increasing f_X also increases the redshift at which $\bar{T}_K \gg T_\gamma$, so that gas temperature fluctuations become irrelevant for the 21 cm signal. Additionally, small values for f_X will increase the contribution of other heating mechanisms such as shock heating. All of this suggests that measuring the time evolution of the 21 cm signal (as in Figure 5.11 for example) would enable f_X to be constrained. Unfortunately, until these observations are made it is difficult to predict the thermal history before reionization with any certainty.

We have shown that the 21 cm signal at high z will contain significantly more structure than has previously been considered. Temperature fluctuations produce an interesting interplay with other sources of 21 cm anisotropy as $\beta_T < 0$ when $T_K < T_\gamma$. Furthermore, for reasonable heating scenarios, the effect of temperature fluctuations persist well into the regime that will be probed by second generation low-frequency arrays, such as the SKA. Thus, prospects for probing the thermal history before reionization via observations of the redshifted 21 cm line seem promising.

Acknowledgments

This work was supported at Caltech in part by DoE DE-FG03-92-ER40701. JRP would like to thank Miguel Morales for useful discussions. SRF thanks the Tapir group at Caltech for hospitality while this work was completed. We would also like to thank the anonymous referee for many useful comments that have helped improve the clarity of the paper.

Chapter 6

Galaxy surveys, inhomogeneous reionization, and dark energy

We examine the effect of inhomogeneous reionization on the galaxy power spectrum and the consequences for probing dark energy. To model feedback during reionization, we apply an *ansatz* setting the galaxy overdensity proportional to the underlying ionization field. Thus, inhomogeneous reionization may leave an imprint in the galaxy power spectrum. We evolve this imprint to low redshift and use the Fisher-matrix formalism to assess the effect on parameter estimation. We show that a combination of low- ($z = 0.3$) and high- ($z = 3$) redshift galaxy surveys can constrain the size of cosmological HII regions during reionization. This imprint can also cause confusion when using baryon oscillations or other features of the galaxy power spectrum to probe the dark energy. We show that when bubbles are large, and hence detectable, our ability to constrain w can be degraded by up to 50%. When bubbles are small, the imprint has little or no effect on measuring dark-energy parameters.

Originally published as Pritchard, Furlanetto, and Kamionkowski, MNRAS, **374**, 159 (2007).

6.1 Introduction

During the epoch of reionization, groups of star-forming regions generate significant numbers of ionizing photons, which may lead to HII regions many Mpc in size (Furlanetto, Zaldarriaga & Hernquist, 2004). These ionized bubbles grow as further structure forms, eventually merging and causing full reionization of the intergalactic medium (IGM). Conditions within these HII regions may be significantly different than in the surrounding neutral IGM. For example, the temperature in these HII regions will be raised by photoionization heating, which is known to suppress star formation in low-mass haloes (Rees, 1986; Efsthathiou, 1992; Thoul & Weinberg, 1996; Kitayama & Ikeuchi, 2000; Dijkstra et al., 2004).

Also, the ionizing flux generates more free electrons, which affects the abundance of molecular hydrogen (Oh & Haiman, 2002), an important coolant. These, and other feedback mechanisms, will affect the fraction of baryons that condense in haloes, and in turn modify the number density of directly observable galaxies (Barkana & Loeb, 2001). This suppression will be inherently inhomogeneous, as highly biased regions will ionize first (Babich & Loeb, 2006). Understanding the detailed effects of feedback is one of the major remaining challenges in understanding galaxy formation.

Galaxies formed during reionization will be low in mass and faint by comparison to those from later generations of galaxy formation. This, along with absorption along the line of sight by the IGM, makes it difficult for existing telescope facilities to detect large numbers of early galaxies directly. However, a number of large galaxy surveys now exist that probe the distribution of galaxies in the lower-redshift regime (Efstathiou et al., 2002; Seljak et al., 2005). These surveys focus on high-mass luminous objects; e.g., luminous red galaxies (LRGs) in SDSS (Eisenstein et al., 2005), which are easily detected. The abundance of such objects will depend in a non-trivial way upon the number of low-mass progenitors, especially upon the amount of condensed gas available for mergers. Thus, these late-forming galaxies will, indirectly, be affected by the efficiency of galaxy formation during reionization. Motivated by these arguments, we consider the possibility that large galaxy surveys in the low-redshift Universe may be used to probe inhomogeneous reionization through its feedback on early galaxy formation. The many uncertainties remaining in our understanding of galaxy formation make it difficult to develop a rigorous formalism for this imprint, and motivate a simpler, hopefully more robust, approach.

Besides the possibility of detecting reionization, its imprint in the galaxy power spectrum may act as a source of noise when probing cosmology. Modern observations of the cosmic microwave background (CMB) (de Bernardis et al., 2000; Halverson et al., 2002; Mason et al., 2003; Benoît et al., 2003; Goldstein et al., 2003; Spergel et al., 2003) have greatly extended our knowledge of cosmological parameters. One result has been the realisation that $\sim 70\%$ of the Universe is composed of an unknown form of energy that generates the accelerated expansion seen in SN Ia observations (Riess et al., 1998; Perlmutter et al., 1999). This is one of the most puzzling discoveries of our times, and it is hoped that future observations in the fields of SN Ia (Riess et al., 1998, 2004; Perlmutter et al., 1999), weak lensing (Hoekstra et al., 2005), and galaxy surveys (Seo & Eisenstein, 2003; Blake

& Glazebrook, 2003) will constrain the time evolution of the dark energy giving clues as to its nature. For this reason, in this paper, we will focus on how reionization may affect estimates of dark-energy parameters.

Large galaxy surveys contribute information on the dark energy in two main ways. First the form of the matter power spectrum, probed by galaxy surveys via the proxy of the galaxy power spectrum, depends on different parameter combinations than the CMB, breaking many of the parameter degeneracies (Eisenstein, Hu & Tegmark, 1999). Second, the pre-recombination oscillation of the photon-baryon fluid leaves an imprint in the matter power spectrum, which may be used as a standard ruler to determine the angular diameter distance $D_A(z)$ as a function of redshift z (Seo & Eisenstein, 2003; Blake & Glazebrook, 2003). These baryon oscillations have now been detected (Cole et al., 2005; Eisenstein et al., 2005) by both 2dF and SDSS. If the imprint in the galaxy power spectrum from patchy reionization can mimic or conceal any feature of the galaxy power spectrum from density fluctuations, then our ability to constrain dark energy using galaxy surveys will be degraded.

In this paper, we explore the possible consequences of this environmental dependence on the galaxy power spectrum. The process of galaxy formation is still only poorly understood and so a detailed analysis of feedback is inappropriate. Instead we choose to link galaxy formation to the neutral fraction by a simple *ansatz*, by which we hope to bring out the underlying behaviour, leaving the details for a later age. In keeping with this “simple is best” ideology, we choose to model the variation in neutral fraction using an analogue of the halo model (Cooray & Sheth, 2002). With this approach we hope to phrase the problem in a general fashion, avoiding detailed assumptions about the reionization history. To address these questions in a quantitative fashion, we employ the Fisher-matrix formalism (Jungman et al., 1996b,a; Tegmark et al., 1997). This allows us to convert a theoretical model into predictions for the parameter constraints attainable by imagined experiments.

The outline of this paper is as follows. In §6.2, we detail the form of the ionization power spectrum and describe our simple *ansatz* relating it to galaxy formation. Then, in §6.3, we bring the two together describing the complete model galaxy power spectrum, including the effects of redshift distortions and the Alcock-Paczynski effect. In §6.4 we outline the Fisher-matrix formalism. Having set out our model, in §6.5 we discuss the possibility of detecting reionization using galaxy surveys. This is then expanded to consider the implications for

dark-energy constraints in §6.6. Finally, in §6.7 we summarise our conclusions.

6.2 Bubble model

We wish to relate the overdensity of galaxies to the ionization fraction within a given region. To do this, we make the simple *ansatz* that there is a component to the galaxy power spectrum which linearly traces the ionized fraction. Hence, we may write the number density of galaxies $n(\mathbf{r})$ at position \mathbf{r} as

$$n(\mathbf{r}) = \bar{n}[1 + \delta_{\text{gal}}(\mathbf{r}) + \delta_{\text{bub}}(\mathbf{r})], \quad (6.1)$$

where \bar{n} is the mean number density of galaxies, $\delta_{\text{gal}}(\mathbf{r}) = b\delta(\mathbf{r})$ assumes galaxies trace the underlying dark-matter fluctuations δ with bias b , and we calculate the fractional overdensity of galaxies due to an ionization field $x_i(\mathbf{r})$ by $\delta_{\text{bub}}(\mathbf{r}) = -\epsilon_b x_i(\mathbf{r})$, where ϵ_b parametrizes the strength of the effect.

Writing the number density in the form of equation (6.1) leads to a galaxy power spectrum

$$P(k) = \frac{1}{(1 - \epsilon_b \bar{Q})^2} [P_{\text{gal}}(k) + 2P_{\text{gal,bub}}(k) + P_{\text{bub}}(k)], \quad (6.2)$$

where \bar{Q} is the filling fraction of the bubbles. For simplicity, we choose to neglect the cross-correlation, which will be smaller or comparable in size to the other terms and represents an unnecessary refinement given the simplicity of our toy model. Note the overall rescaling of the power spectrum because the mean galaxy density $\langle n \rangle = \bar{n}(1 - \epsilon_b \bar{Q})$. Typically, $\epsilon_b \bar{Q} \ll 1$ and we can neglect this correction and take

$$P(k) = P_{\text{gal}}(k) + P_{\text{bub}}(k). \quad (6.3)$$

We now need to calculate the bubble power spectrum $P_{\text{bub}}(k)$. In order to phrase the problem as broadly as possible, we eschew detailed assumptions about reionization in favour of a more general approach. In this paper, we choose to associate regions of ionization with “bubbles,” in analogous fashion to the halo model’s association of mass with haloes. Following the halo-model formalism (Cooray & Sheth, 2002), $P_{\text{bub}}(k)$ is given by the sum of two terms, $P_{\text{bub}}(k) = P^{1b}(k) + P^{2b}(k)$, which describe correlations within the same bubble

and between two different bubbles respectively. These terms are given by

$$P^{1b}(k) = \epsilon_b^2 \int dm n(m) \left(\frac{m}{\bar{\rho}}\right)^2 |u(k|m)|^2, \quad (6.4)$$

$$P^{2b}(k) = \epsilon_b^2 \int dm_1 n(m_1) \left(\frac{m_1}{\bar{\rho}}\right) u(k|m_1) \times \int dm_2 n(m_2) \left(\frac{m_2}{\bar{\rho}}\right) u(k|m_2) P_{bb}(k|m_1, m_2), \quad (6.5)$$

where $n(m)$ is the comoving number density of bubbles of mass m , $P_{bb}(k|m_1, m_2)$ is the power spectrum of bubbles of mass m_1 and m_2 , and $u(k|m)$ is the Fourier transform of the bubble ionization profile $u(\mathbf{r}|m)$. With this notation, we may write the volume filling factor of the bubbles as $\bar{Q} = \int dm n(m)(m/\bar{\rho})$, and the bubble volume as $V_{\text{bub}} = m/\bar{\rho} = 4\pi r_{\text{bub}}^3/3$, where r_{bub} is the comoving bubble radius. Throughout this paper, we will assume a top-hat profile $u(\mathbf{r}|m) = \Theta(|\mathbf{r}| - r_{\text{bub}})/V_{\text{bub}}$, for which $u(k|m) = 3j_1(kr_{\text{bub}})/(kr_{\text{bub}})$, where $j_\ell(x)$ is a spherical Bessel function of order ℓ .

If we assume a delta-function size distribution and treat the power spectrum of the bubbles as tracing the dark-matter power spectrum $P_{bb}(k) \approx P_{\delta\delta}(k, z = z_{ri})$, where z_{ri} is the redshift at which the imprint is formed, this reduces to

$$P^{1b}(k) = \epsilon_b^2 \bar{Q} V_{\text{bub}} |u(k|m)|^2, \quad (6.6)$$

$$P^{2b}(k) = \epsilon_b^2 \bar{Q}^2 |u(k|m)|^2 P_{\delta\delta}(k|m). \quad (6.7)$$

In order to keep our model simple, we ignore evolution in the bubble-size distribution. In reality, the relevant bubble sizes will be determined by the period when most baryons condense, an extended process that will average over the evolution of bubble growth. We also ignore the effects of bubble overlap, which is expected to occur for large \bar{Q} and undermines the halo-model approach. Once bubbles begin to overlap, using isolated spheres to model the HII regions will not correctly represent the true size and shape of the ionized regions. To a first approximation though, this effect will give an effective distribution of bubble sizes, and so should not affect our qualitative conclusions. We will take $\bar{Q} = 0.5$ and $z_{ri} = 6$ in what follows, and use $(r_{\text{bub}}, \epsilon_b)$ to parametrize the bubble power spectrum. Note that the

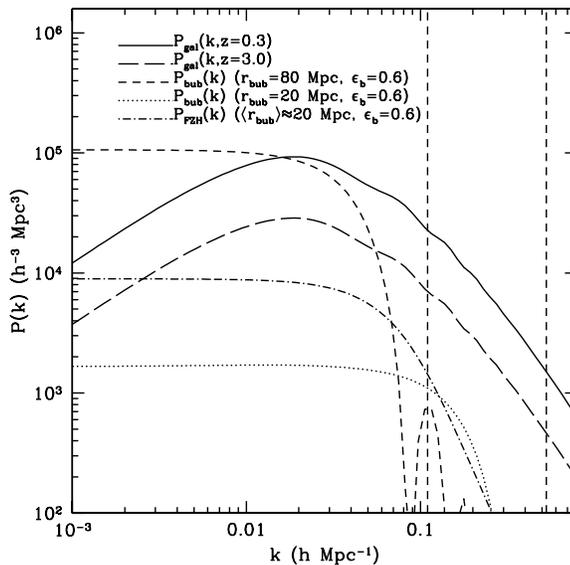


Figure 6.1: Comparison of the galaxy and bubble power spectra. We plot $P_{\text{gal}}(k)$ at two redshifts $z = 0.3$ (solid curve) and $z = 3.0$ (long dashed curve). For each redshift, we plot the non-linear scale: $k_{\text{max}}(z = 0.3) = 0.11h \text{ Mpc}^{-1}$ and $k_{\text{max}}(z = 3.0) = 0.53h \text{ Mpc}^{-1}$ (dashed vertical lines, from left to right). For comparison, we plot $P_{\text{bub}}(k)$ for the parameters ($r_{\text{bub}} = 80 \text{ Mpc}$, $\epsilon_b = 0.6$) (short dashed curve) and ($r_{\text{bub}} = 20 \text{ Mpc}$, $\epsilon_b = 0.6$) (dotted curve). Notice how the latter curve resembles constant white noise in the region $k < 0.1h \text{ Mpc}^{-1}$. The former curve displays a cutoff in power close to the galaxy-power-spectrum peak. Finally, we plot a bubble power spectrum $P_{\text{FZH}}(k)$ (dot-dashed curve) that has been calculated using a bubble size distribution taken from Furlanetto et al. (2004), with $\langle r_{\text{bub}} \rangle \approx 20 \text{ Mpc}$.

two-bubble term is subdominant in the regime that we consider, making the details of z_{ri} , and any biasing of $P_{bb}(k)$ with respect to $P_{\delta\delta}(k)$, unimportant.

Figure 6.1 shows the form of the bubble power spectrum in this model. Note that the power is fairly constant on small k and cuts off sharply on linear scales smaller than the bubble radius. As a simple example of including a smooth bubble size distribution, Figure 6.1 shows P_{bub} calculated using the bubble distribution of Furlanetto et al. (2004), assuming an ionizing efficiency $\zeta = 40$ and $\bar{Q} = 0.83$, which gives a volume averaged bubble size $\langle r_{\text{bub}} \rangle \approx 20 \text{ Mpc}$. The main effects of the distribution in bubble sizes are to smooth out the oscillations seen in the single size model and to decrease the rate at which power decreases on scales below the characteristic bubble size. Having shown that the high- k cutoff occurs even with a smooth distribution of bubble sizes, we will henceforth restrict ourselves to the simpler, single bubble size case.

The onset of non-linearity limits the scales that a galaxy survey is able to probe. We choose to define this cutoff scale by requiring that the average fluctuation on a scale R satisfies $\sigma(R) \leq 0.5$ for $R = \pi/(2k_{\max})$ (Seo & Eisenstein, 2003). This cutoff can lead to a degeneracy between the bubble spectrum and the constant shot-noise expected on large scales from non-Gaussian clustering of the galaxies (Seljak, 2000). We see in Figure 6.1 that on large scales the bubble power spectrum becomes constant. If, for a given galaxy survey, r_{bub} is sufficiently small, then the curvature of the bubble spectrum will lie at $k > k_{\max}$, and the bubble spectrum will be indistinguishable from shot-noise. Including galaxy surveys at higher z , where the non-linear scale is smaller, helps break this degeneracy.

We note that, for a random variable with zero mean, we would expect the power spectrum to vanish on large scales. That this does not occur relates to a generic problem of the halo model 1-halo term, which is constant on large scales. In most applications this is masked by a dominant 2-halo term, which does decrease on large scales. However, in our model the 2-bubble term is negligible making this issue obvious. Given that our model predicts a bubble power spectrum that looks like shot-noise on large scales, we must worry both about how the removal of shot-noise will affect our results and how to distinguish the effect of bubbles from shot-noise. As mentioned above, the existence of a cutoff in the bubble power spectrum distinguishes it from shot-noise (although we must observe this cutoff for this to work). When we come to analyse the effect of bubbles on cosmological parameter estimation, we will include a term representing white shot-noise to account for this possible confusion.

Having generated an imprint from patchy reionization, we must evolve it to lower redshift. Our knowledge of how mergers recycle matter from many smaller haloes into fewer more massive haloes is not sufficient to handle this rigorously. Instead, we will consider three cases that ought to bracket the truth. We take r_{bub} to be a constant, fixing the shape of $P_{\text{bub}}(k)$, and then consider how its amplitude varies with time. We will consider three models for this time evolution

$$P_{\text{bub}}(k, z) = \begin{cases} P_{\text{bub}}(k, z = z_{ri}), & \text{(Model A),} \\ P_{\text{bub}}(k, z = 0) \left[\frac{G(z)}{G(z=0)} \right]^2, & \text{(Model B),} \\ P_{\text{bub}}(k, z = z_{ri}) \left[\frac{G(z)}{G(z=z_{ri})} \right]^{-2}, & \text{(Model C).} \end{cases}$$

In Model A, we assume that, once produced, the power spectrum $P_{\text{bub}}(k)$ from bubbles remains constant in time. As the density fluctuations continue to grow this means that P_{bub} becomes less significant at later times. In Model B, we allow $P_{\text{bub}}(k)$ to grow as the square of the linear growth function $G(z)$. Thus in Model B, $P_{\text{bub}}(k)$ remains a constant fraction of the total galaxy power spectrum. It seems unlikely that the bubble imprint would grow in this fashion, but we include this model in order to consider the case where the bubble imprint is equally important at all redshifts. Note that in this model, we choose to normalise the bubble spectrum to the present day. This provides a simple way of restricting $P_{\text{bub}}(k)$ to amplitudes comparable to the density power spectrum. Finally, with Model C, we consider the case where $P_{\text{bub}}(k)$ decreases with time. This will provide an estimate of the worst-case scenario for detecting the bubbles. The time evolution of $P_{\text{bub}}(k)$ is most important when we can compare surveys at different redshifts. In the case of a single redshift survey, any growth can be absorbed into an effective ϵ_b for that survey.

What range of values may our two free parameters ϵ_b and r_{bub} reasonably take? The characteristic size of the bubbles is the easiest question to address. Furlanetto et al. (2004) present a model for bubbles forming around highly biased regions leading to typical sizes of ~ 5 Mpc, when $\bar{Q} = 0.5$ (see also Furlanetto et al., 2006). In contrast, Wyithe & Loeb (2004) use arguments based on light-travel times and cosmic variance to obtain bubble sizes of ~ 60 Mpc at the end of reionization. This latter value can be taken as an upper limit on reasonable bubble sizes, while the former gives a more reasonable estimate of what we might expect. These values are in broad agreement with the results of computer simulation (Iliev et al., 2005; Zahn et al., 2006), which yield sizes ~ 10 Mpc.

The range of ϵ_b begs the question of how exactly to interpret this parameter. We have assumed a linear relation between the ionization fraction of a region and its galaxy overdensity. We can readily see that $n(\mathbf{r}) \geq 0$, which implies a solid upper limit of $\epsilon_b \leq 1$. An alternative approach is to consider the suppression of galaxy formation in haloes of low mass. Simulations at low redshift ($z < 3$) (Thoul & Weinberg, 1996; Kitayama & Ikeuchi, 2000) indicate significant suppression of galaxy formation in haloes with circular velocities $V_{\text{circ}} \leq 50 \text{ km s}^{-1}$. At higher z , photoionization is less effective due to the decreased cooling time, decreased UV flux, increased self-shielding from the higher densities, and collapse beginning before any UV background can be generated (Dijkstra et al., 2004). In this case, Dijkstra et al. (2004) find that only haloes with $V_{\text{circ}} \leq 20 \text{ km s}^{-1}$ suffer significantly reduced

condensation of baryons. To estimate the mass fraction in galaxies affected by photoionization feedback, we take this latter value and apply it to the Press-Schechter distribution (Press & Schechter, 1974) as a low-mass cutoff below which no galaxies form. This gives an estimate of the decrement in galaxies due to photoionization feedback,

$$\epsilon_b \approx \bar{\Delta}_g \equiv \left[\frac{F(M > M_{\text{feedback}}) - F(M > M_{\text{cool}})}{F(M > M_{\text{cool}})} \right], \quad (6.8)$$

where $F(M)$ is the fraction of mass in haloes of mass greater than M , M_{feedback} is the mass corresponding to $V_{\text{circ}} = 20 \text{ km s}^{-1}$, and M_{cool} is the mass corresponding to the virial temperature $T_{\text{vir}} \approx 10^4 \text{ K}$ needed for effective cooling by atomic hydrogen. Evaluating equation (6.8) gives $\epsilon_b \approx 0.18$ at $z = 10$ and $\epsilon_b \approx 0.10$ at $z = 6$, which gives an indication of sensible values. Once a galaxy grows large enough, gravity will overcome feedback of this form and damp this effect. Thus, these numbers represent an effective upper limit in the most plausible model.

6.3 Galaxy power spectrum

In constructing our galaxy power spectrum, we follow Seo & Eisenstein (2003). Incorporating the effects of bias, linear redshift distortions (Kaiser, 1987), and linear growth, the galaxy power spectrum takes the form

$$P_{\text{gal}}(k, \mu, z) = \left[\frac{G(z)}{G(z=0)} \right]^2 b^2 (1 + \beta \mu^2)^2 P_{\delta}(k, z=0), \quad (6.9)$$

where $P_{\delta}(k, z=0)$ is the power spectrum of the dark matter at the present day, $b = \Omega_m(z)^{0.6}/\beta$ is the bias, and $\mu^2 = k_{\parallel}^2/k^2$ is the direction cosine between the Fourier-mode wavenumber and the line of sight. We define the redshift-distortion parameter β in terms of $\sigma_{8,g}$ and σ_8 , the fluctuations in galaxies and dark matter, respectively, smoothed on scales of $8h^{-1} \text{ Mpc}$, by the relation,

$$\sigma_{8,g} = \sigma_8 b \sqrt{1 + 2\beta/3 + \beta^2/5}. \quad (6.10)$$

In order to calculate the linear growth factor $G(z)$, we integrate the perturbation equation,

$$\ddot{G}(t) + 2H\dot{G}(t) - 4\pi G\rho_m G(t) = 0, \quad (6.11)$$

with

$$\frac{H^2}{H_0^2} = \Omega_m(1+z)^3 + (1 - \Omega_m - \Omega_X)(1+z)^2 + \Omega_X(z), \quad (6.12)$$

and where the energy density in dark energy is given by

$$\Omega_X(z) = \Omega_X \exp\left(3 \int_0^z dz' \frac{1+w(z')}{1+z'}\right). \quad (6.13)$$

In the special case of a cosmological constant, the growth factor may be expressed as

$$G(z) = \frac{5}{2}\Omega_m \frac{H(z)}{H_0} \int_\infty^z \frac{1+z'}{[H(z')/H_0]^3} dz', \quad (6.14)$$

but for a general dark-energy model where $w(z) \neq -1$, the full numerical integration is necessary (Wang & Steinhardt, 1998; Weinberg & Kamionkowski, 2003).

Figure 6.1 shows $P_{\text{gal}}(k, \mu, z)$ averaged over angle and evaluated at $z = 0.3$ and $z = 3$. It displays a clear peak at $k \approx 0.02h \text{ Mpc}^{-1}$, corresponding to the scale of matter-radiation equality, and visible baryon oscillations on smaller scales. These features arise from the acoustic oscillation of the baryon-photon fluid during the period of tight coupling before recombination. The sound speed, which governs the peak positions, is well measured from the CMB. Consequently, the baryon oscillations may be used as a standard ruler, allowing a direct measurement of the angular-diameter distance. These features have now been detected in galaxy surveys (Cole et al., 2005; Eisenstein et al., 2005), and their use in probing the dark energy is well known (Seo & Eisenstein, 2003; Blake & Glazebrook, 2003).

When converting the observed redshift and angular position of galaxies into linear space, we must assume a particular cosmology. If this reference cosmology is different from the true cosmology, then we will introduce distortions into the inferred distribution of galaxies. This is the Alcock-Paczynski (AP) effect (Alcock & Paczynski, 1979) and is essentially a cosmological redshift distortion. We may express the power spectrum inferred by our

observations in terms of the true power spectrum $P^{\text{tr}}(k^{\text{tr}}, \mu^{\text{tr}})$ by

$$P_{\text{obs}}(k, \mu) = \frac{D_A^2(z)H^{\text{tr}}(z)}{D_A^{\text{tr}2}(z)H(z)} P^{\text{tr}}(k^{\text{tr}}, \mu^{\text{tr}}), \quad (6.15)$$

where H and D_A are calculated using the reference cosmology, and H^{tr} and D_A^{tr} with the true cosmology. We write the components of a Fourier wavevector parallel and perpendicular to the line of sight as $k_{\parallel}^{\text{tr}} = (H^{\text{tr}}/H)k_{\parallel}$ and $k_{\perp}^{\text{tr}} = (D_A/D_A^{\text{tr}})k_{\perp}$. The information contained in the AP effect can be useful in probing the evolution of the dark energy, so we include it in this analysis.

The final observed galaxy power spectrum incorporates all of the effects that we have discussed before and takes the form

$$P_{\text{obs}}(k, \mu) = \frac{D_A^2(z)H^{\text{tr}}(z)}{D_A^{\text{tr}2}(z)H(z)} [P_{\text{gal}}(k^{\text{tr}}, \mu^{\text{tr}}) + P_{\text{bub}}(k^{\text{tr}})] + P_{\text{shot}}, \quad (6.16)$$

where P_{shot} is residual shot-noise from non-Gaussian clustering of galaxies (Seljak, 2000), which we treat as a constant white-noise term.

6.4 Fisher matrix

To quantitatively constrain the effect of bubbles on the galaxy power spectrum, we turn to the Fisher matrix. This formalism allows us to estimate the uncertainties on a set of model parameters $\Theta = (\theta_1, \theta_2, \dots, \theta_N)$ given some data set. We define the Fisher matrix (Tegmark et al., 1997)

$$F_{ij} \equiv - \left\langle \frac{\partial^2 \log L}{\partial \theta_i \partial \theta_j} \right\rangle \Big|_{\Theta_0}, \quad (6.17)$$

where L is the likelihood function describing the probability distribution of the parameters and Θ_0 is the place in parameter space where the Fisher matrix is evaluated, typically the point of maximum likelihood. Given the Fisher matrix, the Cramer-Rao inequality states that the minimum uncertainty on a parameter θ_i is given by $\Delta\theta_i \geq (F^{-1})_{ii}^{1/2}$. This estimate of the uncertainty will be reliable provided that Θ_0 is near to the true values of the parameters.

To evaluate F_{ij} , we need to specify a model, which determines the dependence of the

likelihood function on Θ , and a point in parameter space where we wish to determine parameter uncertainties. In the case that the model parameters are Gaussian distributed, the Fisher matrix takes the form

$$F_{\alpha\beta} = \frac{1}{2} \text{tr}(C^{-1} C_{,\alpha} C^{-1} C_{,\beta}) + \frac{\partial\mu}{\partial\theta_\alpha} C^{-1} \frac{\partial\mu}{\partial\theta_\beta}, \quad (6.18)$$

where C is the covariance matrix for the data, and μ is the data's mean. This will be a good approximation in the case of both CMB observations and galaxy survey. Note that, for our purposes, we will need to combine information from both the CMB and galaxy surveys. When used together these data sets break many degeneracies that are present when they are used alone. Let us consider the Fisher matrix from each of these in turn.

A CMB experiment may be characterised by a beam size θ_{beam} and sensitivities to temperature σ_T and polarization σ_P . Given these quantities, the Fisher matrix is given by (Jungman et al., 1996b,a; Kamionkowski et al., 1997; Zaldarriaga & Seljak, 1997)

$$F_{\alpha\beta}^{CMB} = \sum_{\ell} \sum_{X,Y} \frac{\partial C_{\ell}^X}{\partial\theta_\alpha} (\text{Cov}_{\ell})_{XY}^{-1} \frac{\partial C_{\ell}^Y}{\partial\theta_\beta}, \quad (6.19)$$

where C_{ℓ}^X is the power in the ℓ th multipole for $X = T, E, B$, and C , the temperature, E-mode polarization, B-mode polarization, and TE cross-correlation respectively. The elements of the covariance matrix Cov_{ℓ} between the various power spectra are (Kamionkowski et al., 1997; Zaldarriaga & Seljak, 1997)

$$\begin{aligned}
(\text{Cov}_\ell)_{\text{TT}} &= \frac{2}{(2\ell+1)f_{\text{sky}}} (C_{T\ell} + w_T^{-1} B_\ell^{-2})^2, \\
(\text{Cov}_\ell)_{\text{EE}} &= \frac{2}{(2\ell+1)f_{\text{sky}}} (C_{E\ell} + w_P^{-1} B_\ell^{-2})^2, \\
(\text{Cov}_\ell)_{\text{BB}} &= \frac{2}{(2\ell+1)f_{\text{sky}}} (C_{B\ell} + w_P^{-1} B_\ell^{-2})^2, \\
(\text{Cov}_\ell)_{\text{CC}} &= \frac{1}{(2\ell+1)f_{\text{sky}}} [C_{C\ell}^2 + (C_{T\ell} + w_T^{-1} B_\ell^{-2}) \\
&\quad \times (C_{T\ell} + w_T^{-1} B_\ell^{-2})], \\
(\text{Cov}_\ell)_{\text{TE}} &= \frac{2}{(2\ell+1)f_{\text{sky}}} C_{C\ell}^2, \\
(\text{Cov}_\ell)_{\text{TC}} &= \frac{2}{(2\ell+1)f_{\text{sky}}} C_{C\ell} (C_{T\ell} + w_T^{-1} B_\ell^{-2}), \\
(\text{Cov}_\ell)_{\text{EC}} &= \frac{2}{(2\ell+1)f_{\text{sky}}} C_{C\ell} (C_{E\ell} + w_P^{-1} B_\ell^{-2}), \\
(\text{Cov}_\ell)_{\text{TB}} &= (\text{Cov}_\ell)_{\text{EB}} = (\text{Cov}_\ell)_{\text{CB}} = 0.
\end{aligned} \tag{6.20}$$

Here B_ℓ^2 is the beam window function, assumed Gaussian with $B_\ell^2 = \exp[-\ell(\ell+1)\theta_{\text{beam}}^2/8 \ln 2]$, where θ_{beam} is the full-width, half-maximum (FWHM) of the beam in radians. Also, w_T and w_P are the inverse square of the detector noise for temperature and polarization, respectively. For multiple frequency channels we replace $w_T B_\ell^2$ with the sum of this quantity for each of the channels.

Moving now to galaxy surveys, we may write the appropriate Fisher matrix as (Tegmark, 1997)

$$F_{\alpha\beta}^{\text{GAL}} = \int_0^{k_{\text{max}}} \frac{\partial \ln P(\mathbf{k})}{\partial \theta_\alpha} \frac{\partial \ln P(\mathbf{k})}{\partial \theta_\beta} V_{\text{eff}}(\mathbf{k}) \frac{d^3 \mathbf{k}}{2(2\pi)^3}, \tag{6.21}$$

where the derivatives are evaluated using the cosmological parameters of the fiducial model and V_{eff} is the effective volume of the survey, given by

$$\begin{aligned}
V_{\text{eff}}(k, \mu) &= \int d^3 \mathbf{r} \left[\frac{n(\mathbf{r})P(k, \mu)}{n(\mathbf{r})P(k, \mu) + 1} \right]^2 \\
&= \left[\frac{\bar{n}P(k, \mu)}{\bar{n}P(k, \mu) + 1} \right]^2 V_{\text{survey}}.
\end{aligned} \tag{6.22}$$

Here the galaxy survey is parametrized by the survey volume V_{survey} and the galaxy density $n(\mathbf{r})$, which in the last equality we assume to be uniform \bar{n} . In addition, we must specify k_{max} , a cutoff on small scales to avoid the effects of non-linearity. We choose to define this

Table 6.1: Specification for CMB experiments

Experiment	Frequency	θ_{beam}	σ_T	σ_P
WMAP	40	28.2	17.2	24.4
	60	21.0	30.0	42.6
	90	12.6	49.9	70.7
Planck	143	8.0	5.2	10.8
	217	5.5	11.7	24.3

Notes: Frequencies are in GHz. Beam size θ_{beam} is the FWHM in arcsec. Sensitivities σ_T and σ_P are in μK per FWHM beam, $w = (\theta_{\text{beam}}\sigma)^{-2}$. Taken from Eisenstein et al. (1999).

Table 6.2: Specification for galaxy surveys

Survey	z	V_{survey} ($h^{-3}\text{Gpc}^3$)	\bar{n} ($h^3\text{Mpc}^{-3}$)	k_{max} ($h\text{Mpc}^{-1}$)	$\sigma_{8,g}$
SDSS	0.3	1.0	10^{-4}	0.11	1.8
S2	3.0	0.50	10^{-3}	0.53	1.0

Notes: Taken from Seo & Eisenstein (2003).

cutoff scale by the criterion $\sigma(R) \leq 0.5$ for $R = \pi/(2k_{\text{max}})$ (Seo & Eisenstein, 2003).

To apply the above framework, we need a theory relating the observables C_l^X and $P(k, \mu)$ to the parameters. For the CMB, this is standard, while in the case of the galaxy surveys we use equation (6.16), which arose from our discussion in §6.2 and §6.3. Using these models, we calculate the Fisher matrices for individual galaxy surveys and our CMB experiment, and then combine them

$$F_{\alpha\beta}^{TOT} = F_{\alpha\beta}^{CMB} + \sum_i F_{\alpha\beta}^{GAL,i}, \quad (6.23)$$

where i labels the different galaxy surveys. This total Fisher matrix is then inverted to get parameter error predictions.

For this calculation, we need the specifications of our experiments. These are given in Tables 6.1 and 6.2. We consider two galaxy surveys. The first uses parameters corresponding to the SDSS LRG survey, which is currently underway. The second is a hypothetical survey at $z = 3$, based upon a survey of Lyman break galaxies (Seo & Eisenstein, 2003).

Finally, we must decide upon our choice of parameter space. We specify our cosmology using seven parameters describing the total matter fraction $\Omega_m h^2$, Ω_m , the baryon fraction $\Omega_b h^2$, the inflationary amplitude A_S^2 , scalar spectral index n_s , optical depth to last scattering

τ , and the tensor-scalar ratio T/S ; each galaxy survey is described by five parameters ($\log H, \log D_A, \log G, \log \beta, P_{shot}$); to these, we add two parameters ($\epsilon_b, r_{\text{bub}}$) to describe our bubble model (we choose to set $\bar{Q} = 0.5$ and $z_{ri} = 6$). In choosing these parameters, we are following Seo & Eisenstein (2003). We treat all of the above parameters as being independent and then extract information about the dark energy from our uncertainties on ($\log H, \log D_A$) from each survey. We choose to parametrize the dark energy using three parameters (Ω_X, w_0, w_1), taking the dark-energy equation-of-state parameter to be $w(z) = w_0 + w_1 z$. In deciding on our fiducial values, we follow the results of WMAP (Spergel et al., 2003) for the cosmological parameters. These are broadly consistent with the updated results of Spergel et al. (2006), except for the decreased Ω_m and τ . Evaluating the Fisher matrix at these different best-fitting parameters modifies our constraints only slightly. The bubble parameters are highly uncertain, and so we choose to explore a large parameter space.

6.5 Possibility of detecting bubbles

Now that we have established a theoretical framework, we wish to determine whether the imprint can be detected using the specified galaxy surveys. Our null hypothesis is that there is no imprint, and we assume that a detection requires that we can distinguish both ϵ_b and r_{bub} from zero at approximately the 2σ level; i.e., we require both $\epsilon_b > 2\sigma_{\epsilon_b}$ and $r_{\text{bub}} > 2\sigma_{r_{\text{bub}}}$. Throughout, we assume the inclusion of CMB information at the level of Planck. Less precise CMB data will relax constraints on cosmological parameters, causing parameter degeneracies to decrease the sensitivity of the galaxy survey to the bubble imprint.

Figures 6.2, 6.3, and 6.4 show contour plots for models A, B, and C, denoting regions of parameter space where our surveys are able to make a detection. We shade the region of the $(r_{\text{bub}}, \epsilon_b)$ plane where a detection can be made by SDSS alone (white), SDSS and S2 combined (grey), and where no detection can be made (black). For models A and C, we consider the region of parameter space with $r_{\text{bub}} \in [5 \text{ Mpc}, 100 \text{ Mpc}]$ and $\epsilon_b \in [0.1, 1]$. For model B, we choose a slightly different normalisation so that $[G(z=0)/G(z=6)]\epsilon_b \in [0.1, 1]$. This makes the range of amplitude of P_{bub} at $z=0$ identical in the region covered in Figures 6.2 and 6.3. Note that $G(z=6)/G(z=0) = 0.18$ in our fiducial cosmology.

First, compare Figures 6.2 and 6.4. In model A, we see that SDSS alone is able to detect

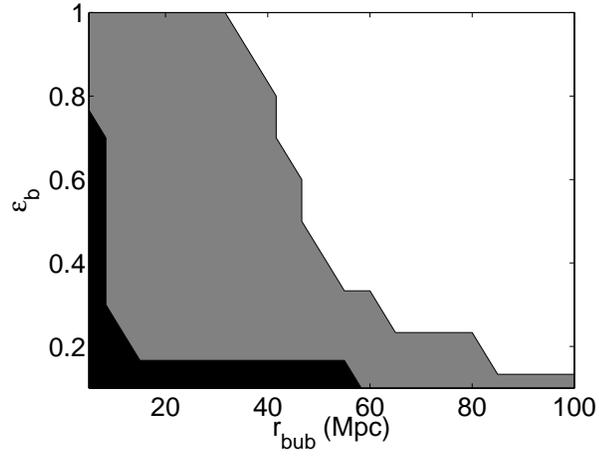


Figure 6.2: Model A: Contour map of detection $\epsilon_b > 2\sigma_{\epsilon_b}$ and $r_{\text{bub}} > 2\sigma_{r_{\text{bub}}}$ in the bubble parameter plane. The white region is detectable by SDSS alone, the grey region is detectable by SDSS+S2, and the black region is undetectable to all surveys. Planck CMB data is assumed in all calculations.

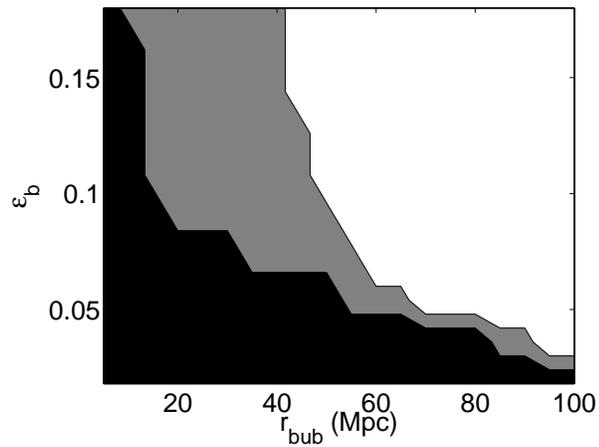


Figure 6.3: Model B: Contour map of detection. As for Figure 6.2. For comparison with other figures, note that $G(z=6)/G(z=0) = 0.18$, so that $[G(z=0)/G(z=6)]\epsilon_b$ lies in the range $[0,1]$.

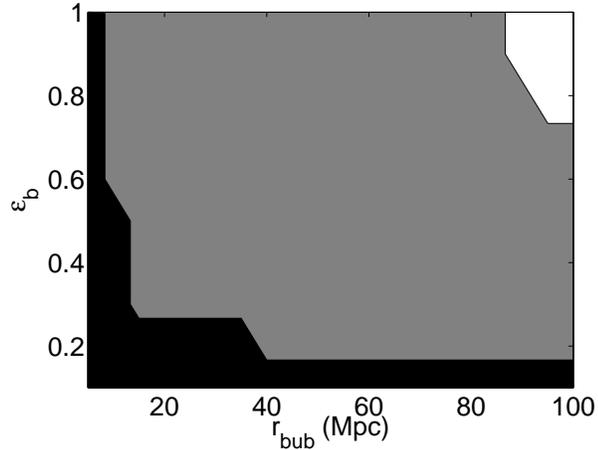


Figure 6.4: Model C: Contour map of detection. As for Figure 6.2. Note the greatly decreased ability of SDSS alone to detect bubbles when compared with Figure 6.2.

bubbles over a wide range of ϵ_b , provided that the bubbles are large ($r_{\text{bub}} > 40$ Mpc). In contrast, when we allow the bubble amplitude to decrease with time, as in Model C, we see that SDSS alone is almost unable to constrain either bubble parameter. In both cases, addition of the S2 survey greatly improves the situation, allowing a wider range of parameter space to be probed. However, even with S2, the theoretically preferred region with $r_{\text{bub}} < 10$ Mpc and $\epsilon_b < 0.2$ (towards the bottom left hand corner) remains unconstrained. The prospects for detection are clearly enhanced by including galaxy surveys at higher redshift.

Figure 6.3, for Model B, shows that growth improves the prospects for probing smaller values of ϵ_b , but makes little difference to our ability to constrain the bubble size. As in model A, SDSS alone can only probe bubbles with $r_{\text{bub}} > 40$ Mpc, and S2 is required to probe smaller scales. Note that when we normalize to the present day, the inclusion of growth in model B reduces the amplitude of the bubble imprint seen by the S2 survey by a factor of $[G(z = 3)/G(z = 0)]^2 \approx 0.1$ over that in model A. This is responsible for the increased region that is undetectable to SDSS+S2 in Figure 6.3. The amplitude of P_{bub} at $z = 0.3$ is very similar in these two models, resulting in the nearly identical contours for SDSS only (when rescaled to account for the different normalisation) in Figures 6.2 and 6.3. The striking differences between the three models indicates the importance of the time evolution of the bubble imprint.

It is worth pausing for a moment to consider where our leverage on the bubble power spectrum originates. When we combine the two surveys, the bulk of the improvement

is coming from the S2 survey alone. This is unsurprising, as the growth of the density fluctuations means the bubble imprint is a more significant contribution to the galaxy power spectrum at early times. Further, if we consider Figure 6.1, we see that for very small bubble sizes, the bubble spectrum begins to resemble white noise over the region probed by the galaxy surveys. This would further complicate detecting the bubble imprint as it could then be confused with residual Poisson shot-noise in the galaxy counts. This problem is greatest at low z where the non-linear scale is larger. Both of these motivate performing this test in galaxy surveys at increasing redshift, ideally at the redshift of reionization, where an $\text{H}\alpha$ survey may be possible.

We conclude that galaxy surveys should be sensitive to the imprint in the galaxy power spectrum left over after reionization. However, detecting this imprint will be difficult unless the characteristic size of HII regions is large ($r_{\text{bub}} > 10 \text{ Mpc}$) and the effects of feedback significant ($\epsilon_b > 0.1$). This should be sufficient to constrain the more extreme models for reionization, but is unlikely to impact more reasonable scenarios. There is significant uncertainty in this prediction stemming from the difficulty in predicting the evolution of the imprint to more recent times.

Currently, the best hope for measuring the size of HII regions during the early stages of reionization lies with upcoming 21cm observations (e.g., LOFAR¹, MWA², or PAST³). Direct imaging of the HII regions is unlikely with the first generation of detectors, but the prospects for statistical detection at $z \leq 10$ are good (Zaldarriaga et al., 2004; Bowman et al., 2006; McQuinn et al., 2005). At higher redshifts, $z > 10$, corresponding to lower frequencies, sky noise increases dramatically making observations more difficult. An imprint directly upon the galaxy power spectrum avoids these technical issues, making possible a complementary measurement. In the event of very early reionization, detection of the imprint discussed in this paper might provide weak constraints on reionization before 21cm experiments reach the desired sensitivity.

¹See <http://www.lofar.org/>.

²See <http://web.haystack.mit.edu/arrays/MWA/>.

³See Pen, Wu & Peterson (2005)

6.6 Implications for dark-energy constraints

Current constraints on dark-energy parameters arise from the combination of high-precision CMB data with information from galaxy surveys. The combination of high- z ($z > 3$) information, long before dark energy becomes dynamically important, with low- z ($z < 3$) information, deep within the dark-energy-dominated regime, serves to break many of the degeneracies that either data set possesses when used alone. Adding in more galaxy surveys at different redshifts further constrains the evolution of the dark energy, allowing constraints on both Ω_X and its equation-of-state parameter $w(z)$. In the previous section, we considered the bubble imprint as a useful signal; in this section, we consider it as a potential source of noise for galaxy surveys. If the bubble power spectrum is able to mimic the effects of dark energy, then it will degrade our ability to constrain dark-energy parameters. Throughout this section, we will consider a dark-energy model with $w_0 = -1$ and $w_1 = 0$. Our numerical results depend upon this choice of model, but the overall picture remains the same when w_0 and w_1 take other values.

Galaxy surveys provide direct constraints on the dark energy through both the baryon oscillations and from the Alcock-Paczynski effect. They also provide indirect constraints in combination with CMB data, as they probe Ω_m independently of $\Omega_m h^2$, the parameter directly probed by the CMB. This allows the CMB indication of flatness $\Omega_k \approx 0$ to constrain Ω_X . The bubble imprint must interfere with one of these measurements to be a source of confusion.

Measurement of the baryon oscillations allow a determination of the angular-diameter distance. Their distinctive oscillatory structure is very different from the smooth structure that we expect from any plausible bubble imprint and so we do not expect there to be any confusion between the two. The inferred peak position, amplitude, and overall shape of the galaxy power spectrum, on the other hand, could be affected by the smooth form of the bubble imprint, making these the most likely points of confusion. Thus, we expect parameters such as Ω_m and n_s to be sensitive to the bubble imprint. This simple picture is modified by inclusion of CMB data, which places tight constraints on many of these parameters, making the effect of the bubble imprint more subtle.

The correlation between the different parameters is indicated in Figure 6.5, for a model with $r_{\text{bub}} = 10 \text{ Mpc}$ and $\epsilon_b = 0.1$. Note that there is a weak correlation between r_{bub} and the

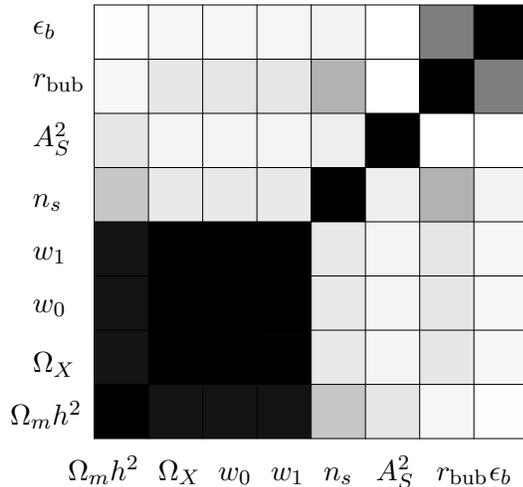


Figure 6.5: Illustration of the reduced covariance matrix. $(\Omega_m h^2, \Omega_X, w_0, w_1, n_s, A_S^2, r_{\text{bub}}, \epsilon_b)$. The model uses $r_{\text{bub}} = 10 \text{ Mpc}$ and $\epsilon_b = 0.1$. Black indicates strong correlation and white indicates little correlation.

dark-energy parameters. A slightly larger covariance is seen between r_{bub} and n_s . At larger values of $(r_{\text{bub}}, \epsilon_b)$, the picture remains unchanged except for a breaking of the degeneracy between r_{bub} and ϵ_b as the cutoff in $P_{\text{bub}}(k)$ on small scales becomes more pronounced.

Before detailing the effect the bubble imprint has on statistical errors, let us consider the possibility of systematic biasing of our best-fitting values, if an existing bubble imprint was ignored in the analysis of data. This will be relevant only in the case that the bubble imprint is not easily detectable, as an obvious imprint would certainly be included in the data analysis. For the case where the bubbles are not detected—i.e., $\epsilon_b < 2\sigma_{\epsilon_b}$ and $r_{\text{bub}} < 2\sigma_{r_{\text{bub}}}$ —we have estimated this systematic offset between the inferred and true parameters, using the Fisher matrix to approximate the full likelihood surface. We find that the offset is significantly smaller than the parameter uncertainty, typically being of order $\sim 0.1\%$. From this, we conclude that failing to include the imprint should not systematically affect parameter estimates in the near future. When galaxy surveys begin to probe cosmological parameters below the percent level this effect will need to be included. We now turn to the effect of the imprint on parameter constraints.

Figures 6.6, 6.7, and 6.8 indicate error contours for w_0 over the $(r_{\text{bub}}, \epsilon_b)$ plane. The same shading scheme is used in all three figures to allow easy comparison. First consider Figure 6.6. We see that the uncertainty on w_0 is maximal for bubble parameters $r_{\text{bub}} \approx 80$ Mpc and $\epsilon_b \approx 0.5$. The form of $P_{\text{bub}}(k)$ is plotted in Figure 6.1 where we see that it cuts off close to the maximum of the density power spectrum. This is consistent with our above statements. We find a maximum uncertainty of $\sigma_{w_0} = 0.48$ in contrast with the uncertainty $\sigma_{w_0} = 0.39$ in the absence of bubbles. This indicates that bubbles can be an important source of noise in attempts to constrain dark energy. However, the large values of $(r_{\text{bub}}, \epsilon_b)$ required for this effect seem theoretically unlikely and from the discussion in §6.5 would allow direct detection of the bubbles. For more reasonable choices of bubble parameters ($r_{\text{bub}} < 10$ Mpc, $\epsilon_b < 0.1$), the uncertainty on w_0 reduces to $\sigma_{w_0} = 0.39$. Thus, the effect of the bubble imprint is likely to be somewhat important in future attempts to constrain dark energy.

Now consider Figures 6.7 and 6.8. The increased uncertainty in w_0 is more pronounced in Model B, where the uncertainty rises as high as $\sigma_{w_0} = 0.62$. Even here, for small values of $(r_{\text{bub}}, \epsilon_b)$, the bubble imprint becomes unimportant and we recover $\sigma_{w_0} = 0.39$, the no-bubble uncertainty. Again this maximal uncertainty occurs close to $r_{\text{bub}} \approx 80$ Mpc and $\epsilon_b \approx 0.5$. The increased value of σ_{w_0} is a consequence of the bubble imprint growing at the same rate as the density fluctuations. Consequently, the overall shape of the total galaxy power spectrum remains constant in time. Thus, combining information at two redshifts provides much less leverage on separating out the two components, leading to larger parameter uncertainties. In model C, the damping of the imprint means that it has much less effect on the dark-energy parameters.

Finally, we note that Figures 6.6, 6.7, and 6.8 display a region of decreased uncertainty in w_0 in the top right hand corner, when the bubbles are large and feedback strong. This is an interesting example of the AP effect. In this region, the bubble power spectrum dominates over the density contribution and the overall shape of the galaxy power spectrum displays a well defined, sharp cutoff. Distortion of this scale by the AP effect places good constraints on the dark energy. Galaxy surveys already show that the galaxy power spectrum closely traces the underlying density field, so this region is ruled out.

Having considered dark-energy parameters, it would seem natural to also consider inflationary parameters; e.g., the tilt n_s and amplitude A_S^2 . For the surveys that we have

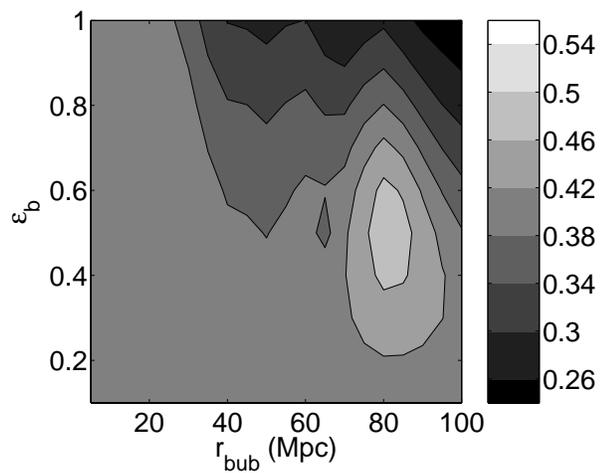


Figure 6.6: Model A: Contour map of errors in w_0 in the bubble parameter plane. We plot contours spanning the range $\sigma_{w_0} = 0.22$ – 0.58 in intervals of 0.04 . The fiducial model takes $w_0 = -1$ and $w_1 = 0$. When no bubbles are present, we find $\sigma_{w_0} = 0.39$.

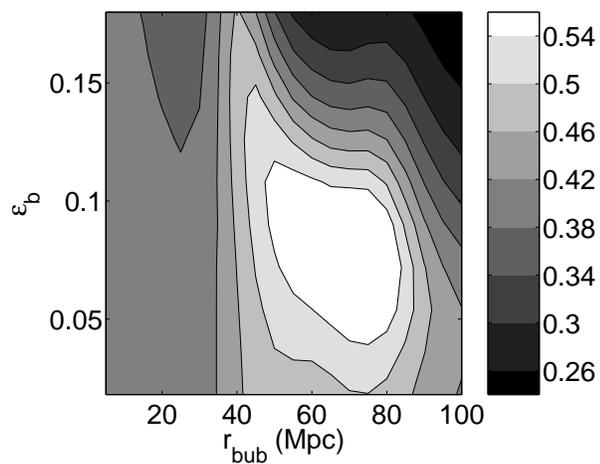


Figure 6.7: Model B: Contour map of errors in w_0 in the bubble parameter plane. As for Figure 6.6.

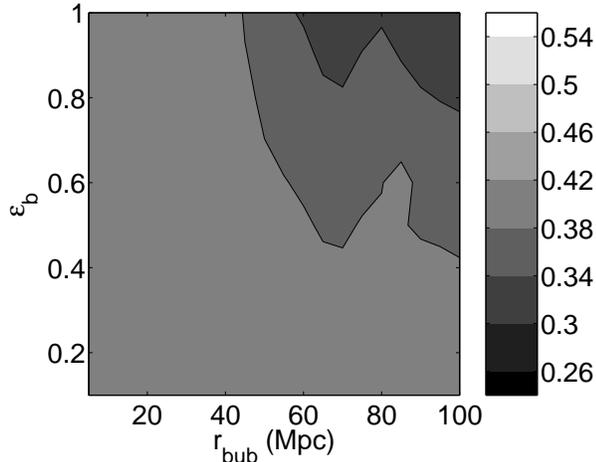


Figure 6.8: Model C: Contour map of errors in w_0 in the bubble parameter plane. As for Figure 6.6.

analysed, inclusion of the bubble power spectrum makes little difference to the uncertainty on these parameters. Essentially, all of the information needed for constraining these quantities is contained within the CMB. In the absence of information on the optical depth τ , or if there are significant tensor modes, galaxy-survey information becomes important in breaking degeneracies. This is not true in the cases that we consider, where CMB-polarization information is well measured. If we were to try to use galaxy-survey data by itself, we would notice increased uncertainty in the tilt n_s .

6.7 Conclusions

In this paper, we have discussed the possibility that patchy reionization may leave an imprint in the distribution of galaxies through its effect on the collapse and cooling of baryons. We considered a simple *ansatz*, linking galaxy number density to the ionization fraction, and used a halo-model approach to calculate the imprint of inhomogeneous ionization on the galaxy power spectrum. We then applied a Fisher-matrix approach to place constraints on the effect of this imprint.

Our calculation shows that detecting the bubble imprint through large galaxy surveys is potentially feasible, but highly dependent upon the details of reionization. We have shown that, for a detection to be possible with upcoming experiments, bubbles must be large ($r_{\text{bub}} > 10 \text{ Mpc}$) and the feedback moderately strong ($\epsilon_b > 0.1$). This suggests that the

most reasonable region of parameter space ($r_{\text{bub}} < 10 \text{ Mpc}$, $\epsilon_b < 0.1$) will not be detected with currently proposed galaxy surveys at $z \leq 3$. Potentially, a $z \approx 6$ galaxy survey might give the additional leverage needed for a concrete detection.

Beyond the possibility of detection, we have considered the effect of the bubble imprint on constraining dark-energy parameters. We find that the distinctive nature of the baryon oscillations helps minimize any degeneracy arising. Only if the characteristic bubble size is $\sim 80 \text{ Mpc}$ does the bubble imprint seriously impact our uncertainty in w_0 . In this case, the bubble power spectrum closely mimics the cutoff of the density power spectrum. This is a region of parameter space where the bubbles should be easily detected. For more-sensible values of $(r_{\text{bub}}, \epsilon_b)$, there is little or no impact on dark-energy constraints. When the bubbles are not detectable, we find that ignoring them in the analysis of galaxy data does not introduce any significant biasing of the best-fitting parameters.

Our approach has emphasised the use of a simple toy model to probe the effect of reionization on the distribution of galaxies. If future galaxy surveys are able to make detections of this signal, it will be important to incorporate more-detailed physics to better constrain the shape and amplitude of the bubble imprint. With our present understanding of reionization, this seems premature.

Future galaxy surveys will greatly add to our knowledge of the distribution of galaxies and the nature of the dark energy. If we are to extract maximum information from these surveys, we must tighten our understanding of the biasing of galaxy formation and the possible effect of reionization on early generations of galaxies.

Acknowledgments

This work was supported at Caltech in part by DoE DE-FG03-92-ER40701 and NASA NNG05GF69G.

Bibliography

Abbott L., Wise M., 1984, *Nuc. Phys. B*, 237, 226

Albrecht A., Steinhardt P. J., 1982, *Phys Rev Lett*, 48, 1220

Alcock C., Paczynski B., 1979, *Nature*, 281, 358

Allison A. C., Dalgarno A., 1969, *ApJ*, 158, 423

Atrio-Barandela F., Silk J., 1994, *Phys Rev D*, 49, 1126

Babich D., Loeb A., 2006, *ApJ*, 640, 1

Bardeen J. M., Steinhardt P. J., Turner M. S., 1983, *Phys Rev D*, 28, 679

Barkana R., Loeb A., 2001, *Physics Reports*, 349, 125

Barkana R., Loeb A., 2005a, *ApJ*, 624, L65

Barkana R., Loeb A., 2005b, *ApJ*, 626, 1

Barkana R., Loeb A., 2005c, *MNRAS*, 363, L36

Bartolo N., Komatsu E., Matarrese S., Riotto A., 2004, *Physics Reports*, 402, 103

Basko M. M., 1981, *Astrophysics*, 17, 69

Bender C., Orszag S., 1978, *Advanced Mathematical Methods for Scientists and Engineers*.
McGraw-Hill Book Company, New York

Benoît A., et al., 2003, *A & A*, 399, L25

Bertschinger E., 1996, in *Cosmology and Large Scale Structure*. Les Houches Session LX,
Elsevier, Amsterdam

- Bharadwaj S., Ali S. S., 2004, MNRAS, 352, 142
- Blake C., Glazebrook K., 2003, ApJ, 594, 665
- Bond J. R., 1996, in *Cosmology and Large Scale Structure. Les Houches Session LX*, Elsevier, Amsterdam
- Bowman J. D., Morales M. F., Hewitt J. N., 2006, ApJ, 638, 20
- Bray I., Fursa D. V., Kheifets A. S., Stelbovics A. T., 2002, J. Phys. B: At. Mol. Opt. Phys., 35, 117
- Breit G., Teller E., 1940, ApJ, 91, 215
- Brunner H., 1988, in *Numerical Analysis 1987. Vol. 170 of Pitman Research Notes in Mathematics*, Longman Scientific and Technical, Harlow, Essex, UK
- Cabella P., Kamionkowski M., 2003, Lectures given at the 2003 Villa Mondragone School of Gravitation and Cosmology, astro-ph/0403392
- Carilli C. L., Rawlings S., 2004, New Astronomy Review, 48, 979
- Chen X., Miralda-Escudé J., 2004, ApJ, 602, 1
- Chen X., Miralda-Escudé J., 2006, ApJ, submitted (astro-ph/0605439)
- Chugai N. N., 1980, Soviet Astronomy Letters, 6, 91
- Chugai N. N., 1987, Astrofizika, 26, 89
- Chuzhoy L., Alvarez M. A., Shapiro P. R., 2006, ApJ, 648, L1
- Chuzhoy L., Shapiro P. R., 2006a, ApJ, submitted (astro-ph/0604483)
- Chuzhoy L., Shapiro P. R., 2006b, ApJ, 651, 1
- Ciardi B., Madau P., 2003, ApJ, 596, 1
- Cole S., et al., 2005, MNRAS, 362, 505
- Cooray A., 2006, Phys Rev D, 73, 103001
- Cooray A., Melchiorri A., Silk J., 2003, Phys. Lett., B554, 1

- Cooray A., Sheth R., 2002, *Physics Reports*, 372, 1
- Crittenden R., 1993, PhD thesis, University of Pennsylvania
- de Bernardis P., et al., 2000, *Nature*, 404, 955
- Debye P., 1909, *Math. Ann.*, 67, 535
- Di Matteo T., Ciardi B., Miniati F., 2004, *MNRAS*, 355, 1053
- Dijkstra M., Haiman Z., Loeb A., 2004, *ApJ*, 613, 646
- Dijkstra M., Haiman Z., Rees M. J., Weinberg D. H., 2004, *ApJ*, 601, 666
- Dodelson S., 2003, *Modern Cosmology*. Academic Press, San Diego, USA
- Efstathiou G., 1992, *MNRAS*, 256, 43P
- Efstathiou G., et al., 2002, *MNRAS*, 330, L29
- Eisenstein D. J., et al., 2005, *ApJ*, 633, 560
- Eisenstein D. J., Hu W., Tegmark M., 1999, *ApJ*, 518, 2
- Fabbri R., Pollock M. D., 1983, *Phys. Lett. B*, 125, 445
- Field G. B., 1958, *Proc. I. R. E.*, 46, 240
- Field G. B., 1959a, *ApJ*, 129, 536
- Field G. B., 1959b, *ApJ*, 129, 551
- Furlanetto S., Furlanetto M., 2006, *MNRAS*, submitted (astro-ph/0608067)
- Furlanetto S. R., 2006, *MNRAS*, 371, 867
- Furlanetto S. R., Loeb A., 2004, *ApJ*, 611, 642
- Furlanetto S. R., McQuinn M., Hernquist L., 2006, *MNRAS*, 365, 115
- Furlanetto S. R., Oh S. P., Briggs F. H., 2006, *Physics Reports*, 433, 181
- Furlanetto S. R., Pritchard J. R., 2006, *MNRAS*, 372, 1093

- Furlanetto S. R., Schaye J., Springel V., Hernquist L., 2005, *ApJ*, 622, 7
- Furlanetto S. R., Sokasian A., Hernquist L., 2004, *MNRAS*, 347, 187
- Furlanetto S. R., Zaldarriaga M., Hernquist L., 2004, *ApJ*, 613, 1
- Gilfanov M., Grimm H.-J., Sunyaev R., 2004, *MNRAS*, 347, L57
- Glover S. C. O., Brand P. W. J. L., 2003, *MNRAS*, 340, 210
- Goldstein J. H., et al., 2003, *ApJ*, 599, 773
- Grachev S. I., 1989, *Astrofizika*, 30, 347
- Gunn J. E., Peterson B. A., 1965, *ApJ*, 142, 1633
- Guth A. H., 1981, *Phys Rev D*, 23, 347
- Guth A. H., Pi S.-Y., 1982, *Phys Rev Lett*, 49, 1110
- Halverson N. W., et al., 2002, *ApJ*, 568, 38
- Hanany S., et al., 2000, *ApJ*, 545, L5
- Hawking S. W., 1982, *Phys. Lett.*, B115, 29
- Henry L. G., 1941, *Proc. Nat. Acad. Sci.*, 26, 50
- Hirata C. M., 2006, *MNRAS*, 367, 259
- Hirata C. M., Seljak U., 2003a, *Phys Rev D*, 67, 043001
- Hirata C. M., Seljak U., 2003b, *Phys Rev D*, 68, 083002
- Hirata C. M., Sigurdson K., 2006, *MNRAS*, submitted (astro-ph/0605071)
- Hoekstra H., et al., 2005, *ApJ*, submitted (astro-ph/0511089)
- Hogan C. J., Rees M. J., 1979, *MNRAS*, 188, 791
- Hu W., Fukugita M., Zaldarriaga M., Tegmark M., 2001, *ApJ*, 549, 669
- Hu W., Hedman M. M., Zaldarriaga M., 2003, *Phys Rev D*, 67, 043004

- Hu W., Okamoto T., 2002, *ApJ*, 574, 566
- Hu W., Sugiyama N., 1995, *ApJ*, 444, 489
- Hu W., Sugiyama N., 1996, *ApJ*, 471, 542
- Hu W., White M., 1997a, *Phys Rev D*, 56, 596
- Hu W., White M., 1997b, *New Astron.*, 2, 323
- Hui L., Haiman Z., 2003, *ApJ*, 596, 9
- Hummer D. G., 1962, *MNRAS*, 125, 21
- Hummer D. G., Rybicki G. B., 1992, *ApJ*, 387, 248
- Iliev I. T., Mellema G., Pen U.-L., Merz H., Shapiro P. R., Alvarez M. A., 2005, *MNRAS*, submitted (astro-ph/0512187)
- Jaffe A. H., Kamionkowski M., Wang L., 2000, *Phys Rev D*, 61, 083501
- Jones B., Wyse R., 1985, *A & A*, 149, 144
- Jungman G., Kamionkowski M., Kosowsky A., Spergel D. N., 1996a, *Phys Rev Lett*, 76, 1007
- Jungman G., Kamionkowski M., Kosowsky A., Spergel D. N., 1996b, *Phys Rev D*, 54, 1332
- Kaiser N., 1987, *MNRAS*, 227, 1
- Kamionkowski M., Kosowsky A., 1998, *Phys Rev D*, 57, 685
- Kamionkowski M., Kosowsky A., 1999, *Ann. Rev. Nucl. Part. Sci.*, 49, 77
- Kamionkowski M., Kosowsky A., Stebbins A., 1997, *Phys Rev Lett*, 78, 2058
- Kamionkowski M., Kosowsky A., Stebbins A., 1997, *Phys Rev D*, 55, 7368
- Kamionkowski M., Spergel D., Sugiyama N., 1994, *ApJ*, 426, L57
- Keating B., et al., 1998, *ApJ*, 495, 580
- Kesden M., Cooray A., Kamionkowski M., 2002, *Phys Rev Lett*, 89, 011304

- Kesden M., Cooray A., Kamionkowski M., 2003, *Phys Rev D*, 67, 123507
- Kitayama T., Ikeuchi S., 2000, *ApJ*, 529, 615
- Knox L., Song Y.-S., 2002, *Phys Rev Lett*, 89, 011303
- Kosowsky A., 1996, *Ann. Phys.*, 246, 49
- Kuhlen M., Madau P., 2005, *MNRAS*, 363, 1069
- Kuhlen M., Madau P., Montgomery R., 2006, *ApJ*, 637, L1
- Lewis A., Challinor A., Turok N., 2002, *Phys Rev D*, 65, 023505
- Lin Y.-T., Wandelt B. D., 2006, *Astropart. Phys.*, 25, 151
- Linde A. D., 1982a, *Phys. Lett.*, B108, 389
- Linde A. D., 1982b, *Phys. Lett.*, B116, 335
- Loeb A., Rybicki G. B., 1999, *ApJ*, 524, 527
- Loeb A., Zaldarriaga M., 2004, *Phys Rev Lett*, 92, 211301
- Lue A., Wang L., Kamionkowski M., 1999, *Phys Rev Lett*, 83, 1506
- Lyth D., Riotto A., 1999, *Physics Reports*, 314, 1L
- Madau P., Meiksin A., Rees M. J., 1997, *ApJ*, 475, 429
- Madau P., Rees M. J., Volonteri M., Haardt F., Oh S. P., 2004, *ApJ*, 604, 484
- Mason B. S., et al., 2003, *ApJ*, 591, 540
- Mather J. C., et al., 1990, *ApJ*, 354, L37
- McDonald P., Miralda-Escudé J., Rauch M., Sargent W. L. W., Barlow T. A., *Cen R.*, 2001, *ApJ*, 562, 52
- McQuinn M., Zahn O., Zaldarriaga M., Hernquist L., Furlanetto S. R., 2005, *ApJ*, submitted (astro-ph/0512263)
- Meiksin A., 2006, *MNRAS*, 370, 2025

- Miller A. D., et al., 1999, *ApJ*, 524, L1
- Miralda-Escudé J., Haehnelt M., Rees M. J., 2000, *ApJ*, 530, 1
- Morales M. F., 2005, *ApJ*, 619, 678
- Morales M. F., Hewitt J., 2004, *ApJ*, 615, 7
- Naoz S., Barkana R., 2005, *MNRAS*, 362, 1047
- Ng K., Speliotopoulos A., 1995, *Phys Rev D*, 52, 2112
- Ng K. L., Ng K.-W., 1996, *ApJ*, 456, 413
- Nusser A., 2005, *MNRAS*, 359, 183
- Oh S. P., 2001, *ApJ*, 553, 499
- Oh S. P., Haiman Z., 2002, *ApJ*, 569, 558
- Oh S. P., Mack K. J., 2003, *MNRAS*, 346, 871
- Ostriker J. P., Gnedin N. Y., 1996, *ApJ*, 472, L63
- Peebles P. J. E., 1993, *Principles of physical cosmology*. Princeton Series in Physics, Princeton, NJ: Princeton University Press, —c1993
- Peebles P. J. E., Yu J. T., 1970, *ApJ*, 162, 815
- Pen U. L., Wu X. P., Peterson J., 2005, *ChJAA*, submitted (astro-ph/0404083)
- Perlmutter S., et al., 1999, *ApJ*, 517, 565
- Polnarev A., 1985, *Sov. Astron.*, 29, 6
- Press W., Teukolsky S., Vetterling W., Flannery B., 1992, *Numerical Recipes in C*. Cambridge University Press, Cambridge, UK
- Press W. H., Schechter P., 1974, *ApJ*, 187, 425
- Pritchard J. R., Furlanetto S. R., 2006, *MNRAS*, 367, 1057
- Pritchard J. R., Furlanetto S. R., 2007, *MNRAS*, 376, 1680

- Pritchard J. R., Furlanetto S. R., Kamionkowski M., 2006, MNRAS, 374, 159
- Pritchard J. R., Kamionkowski M., 2005, Annals of Physics, 318, 2
- Ranalli P., Comastri A., Setti G., 2003, A & A, 399, 39
- Rees M. J., 1986, MNRAS, 218, 25
- Ricotti M., Ostriker J. P., 2004, MNRAS, 352, 547
- Riess A. G., et al., 1998, AJ, 116, 1009
- Riess A. G., et al., 2004, ApJ, 607, 665
- Rubakov V. A., Sazhin M., Veryaskin A., 1982, Phys. Lett. B, 115, 189
- Rudnick P., 1935, Phys. Rev., 114, 114
- Rybicki G. B., 2006, ApJ, 647, 709
- Rybicki G. B., dell'Antonio I. P., 1994, ApJ, 427, 603
- Santos M. G., Cooray A., 2006, Phys Rev D, 74, 083517
- Santos M. G., Cooray A., Knox L., 2005, ApJ, 625, 575
- Schaye J., Theuns T., Rauch M., Efstathiou G., Sargent W. L. W., 2000, MNRAS, 318, 817
- Scott D., Rees M. J., 1990, MNRAS, 247, 510
- Seager S., Sasselov D. D., Scott D., 1999, ApJ, 523, L1
- Seljak U., 2000, MNRAS, 318, 203
- Seljak U., et al., 2005, Phys Rev D, 71, 103515
- Seljak U., Hirata C. M., 2004, Phys Rev D, 69, 043005
- Seljak U., Zaldarriaga M., 1996, ApJ, 469, 437
- Seljak U., Zaldarriaga M., 1997, Phys Rev Lett, 78, 2054
- Seljak U., Zaldarriaga M., 1999, Phys Rev Lett, 82, 2636

- Seo H.-J., Eisenstein D. J., 2003, *ApJ*, 598, 720
- Sheth R. K., Tormen G., 1999, *MNRAS*, 308, 119
- Shull J. M., 1979, *ApJ*, 234, 761
- Shull J. M., van Steenberg M. E., 1985, *ApJ*, 298, 268
- Smith T. L., Kamionkowski M., Cooray A., 2006, *Phys Rev D*, 73, 023504
- Sobelman I. I., 1972, *Introduction to the Theory of Atomic Spectra*. Pergamon Press, Oxford, UK
- Spergel D. N., et al., 2003, *ApJS*, 148, 175
- Spergel D. N., et al., 2006, *ApJ*, submitted (astro-ph/0603449)
- Starobinsky A., 1985, *Sov. Astron. Lett.*, 11, 133
- Starobinsky A. A., 1982, *Phys. Lett.*, B117, 335
- Sunyaev R., Zeldovich Y., 1970, *Astrophys. & Space Sci.*, 7, 3
- Tegmark M., 1997, *Phys Rev Lett*, 79, 3806
- Tegmark M., Taylor A. N., Heavens A. F., 1997, *ApJ*, 480, 22
- Theuns T., Schaye J., Zaroubi S., Kim T.-S., Tzanavaris P., Carswell B., 2002, *ApJ*, 567, L103
- Thoul A. A., Weinberg D. H., 1996, *ApJ*, 465, 608
- Turner M., White M., Lidsey J., 1993, *Phys Rev D*, 48, 4613
- Unno W., 1952, *PASJ*, 3, 158
- Venkatesan A., Giroux M. L., Shull J. M., 2001, *ApJ*, 563, 1
- Verner D. A., Ferland G. J., Korista K. T., Yakovlev D. G., 1996, *ApJ*, 465, 487
- Wang L., Steinhardt P. J., 1998, *ApJ*, 508, 483
- Wang X., Hu W., 2006, *ApJ*, 643, 585

- Wang X., Tegmark M., Zaldarriaga M., 2002, Phys Rev D, 65, 123001
- Wang Y., 1996, Phys Rev D, 53, 639
- Weinberg N. N., Kamionkowski M., 2003, MNRAS, 341, 251
- Weinberg S., 2004, Phys Rev D, 69, 023503
- Wouthuysen S. A., 1952, AJ, 57, 31
- Wyithe J. S. B., Loeb A., 2004, Nature, 432, 194
- Zahn O., Lidz A., McQuinn M., Dutta S., Hernquist L., Zaldarriaga M., Furlanetto S. R.,
2006, ApJ, submitted (astro-ph/0604177)
- Zaldarriaga M., 1997, Phys Rev D, 55, 1822
- Zaldarriaga M., Furlanetto S. R., Hernquist L., 2004, ApJ, 608, 622
- Zaldarriaga M., Harari D., 1995, Phys Rev D, 52, 3276
- Zaldarriaga M., Hui L., Tegmark M., 2001, ApJ, 557, 519
- Zaldarriaga M., Seljak U., 1997, Phys Rev D, 55, 1830
- Zheng W., Kriss G. A., Telfer R. C., Grimes J. P., Davidsen A. F., 1997, ApJ, 475, 469
- Zygelman B., 2005, ApJ, 622, 1356

Appendix A

Numerical evolution of gravitational wave amplitude

A.1 Numerical evolution of gravitational waves with anisotropic stress

In this Appendix, we return to the question of anisotropic stress. In the early Universe, free-streaming neutrinos provide the main source of anisotropic stress. After recombination, photons free stream and can also contribute, though the energy density in radiation is falling fast and the effect is negligible. Working from Eq. (2.11) and standard expressions for the energy density of a distribution of relativistic massless particles, an integro-differential equation describing the evolution of the tensor modes may be derived (Weinberg, 2004),

$$\ddot{h} + 2\frac{\dot{a}}{a}\dot{h} + k^2h = -24f_\nu(\tau) \left(\frac{\dot{a}(\tau)}{a(\tau)}\right)^2 \int_0^\tau K[k(\tau - \tau')] \dot{h}(\tau') d\tau', \quad (\text{A.1})$$

where $f_\nu \equiv \bar{\rho}_\nu/\bar{\rho}$ with $\bar{\rho}$ the unperturbed density, and $K(s)$ is given by

$$K(s) \equiv -\frac{\sin s}{s^3} - \frac{3 \cos s}{s^4} + \frac{3 \sin s}{s^5}. \quad (\text{A.2})$$

The new term acts to damp the amplitude of h and can be seen to have the form of a convolution over the mode's past history of the kernel $K(s)$ and the “velocity” \dot{h} . The linear dependence on f_ν means that the damping term will become negligible in the matter-dominated regime where $f_\nu \propto a^{-1}$. In the radiation-dominated epoch, though, the neutrino and total energy densities scale in the same way, leading to $f_\nu = 0.40523$. This suggests that the damping term will primarily affect those modes that enter the horizon well within

the radiation-dominated epoch. In consequence, it will affect the power spectrum only at high l where these modes are the dominant contributors.

The right-hand side is also damped by the $(\dot{a}/a)^2$ term which scales as τ^{-2} in both the matter- and radiation-dominated epochs.

Numerical integration of Eq. (A.1) is possible after recasting the integro-differential equation as a set of coupled Volterra integral equations (Brunner, 1988). These coupled equations may then be integrated using standard techniques (Press et al., 1992).

Given an integro-differential equation of the form,

$$y^{(r)}(t) = f(t, y(t), \dots, y^{(r-1)}(t)) + \int_0^t K(t, s, y(s), \dots, y^{(r)}(s)) ds, \quad (\text{A.3})$$

and defining $z_k(t) = y^{(k)}(t)$ for $k = 0, \dots, r-1$, we may recast Eq. (A.3) as the set of first-order Volterra integral equations,

$$z_r(t) = \int_0^t K(t, s, y(s), \dots, y^{(r-1)}(s)) ds, \quad (\text{A.4})$$

$$z_{r-1}(t) = y_0^{(r-1)} + \int_0^t (f(s, z_0(s), \dots, z_{r-1}(s)) + z_r(s)) ds, \quad (\text{A.5})$$

$$z_k(t) = y_0^{(k)} + \int_0^t z_{k+1}(s) ds, \quad (\text{A.6})$$

where the y_0^k are the relevant initial conditions.

Making the co-ordinate transformation $t = k\tau$ in Eq. (A.1), we can apply this formalism to obtain

$$z_2(t) = \int_0^t K_n(t-s) z_1(s) ds, \quad (\text{A.7})$$

$$z_1(t) = y_0^1 + \int_0^t \left(-2 \frac{a'(s)}{a(s)} z_1(s) - z_0(s) + z_2(s) \right) ds, \quad (\text{A.8})$$

$$z_0(t) = y_0 + \int_0^t z_1(s) ds, \quad (\text{A.9})$$

where primes denote differentiation with respect to t , and we have defined $z_1 = h'$, $z_0 = h$, and

$$K_n(s) = -24 f_\nu(t) \left(\frac{\dot{a}(t)}{a(t)} \right)^2 K(t-s). \quad (\text{A.10})$$

Fig. 2.2 shows the results of numerical integration of Eq. (A.1). The inclusion of

anisotropic stress damps the wave during the radiation-dominated regime leading to decreased amplitude and a slightly shifted phase.

Of primary interest here is the effect of the anisotropic stress by the time of recombination. It is traditional to calculate a transfer function relating the amplitude and phase of the numerical solution to that of the matter-dominated solution $h_{\text{mat}}(\tau) = 3j_1(k\tau)/k\tau$. At recombination, radiation is still important and this analytic form is a relatively poor approximation. For illustrative purposes, we will numerically calculate the amplitude ratio A and phase shift Ψ between numerical calculations of h with and without the effects of anisotropic stress evaluated at the time of recombination. To calculate A at a given τ , we first calculate Ψ and then numerically fit $h_{\text{stress}}(k\tau)$ with $A h_{\text{no-stress}}(k\tau + \Psi)$ over the period containing τ . This avoids the oscillation that results if we seek to obtain the amplitude ratio by simply dividing $h_{\text{stress}}(k\tau)$ by $h_{\text{no-stress}}(k\tau)$.

The results of this calculation (Fig. A.1) illustrate that anisotropic stress introduces a k -dependent damping asymptoting to a factor of $A \sim 0.81$. The phase shift introduced remains small and reaches a maximum value of $\Psi \sim 0.13$ rad. The amplitude ratio is unity while the gravitational wave remains within the horizon. Around horizon entry it rises slightly above unity before decreasing asymptotically to $A \sim 0.81$. This slight rise is a consequence of the anisotropic stress which, like viscosity acting on a pendulum, slows the initial decrease of the gravitational wave and leads to a lower final amplitude of oscillation. The phase difference between damped and undamped cases grows after horizon entry as a consequence of the slower evolution of the damped wave. It peaks and begins to asymptote to a constant value for modes that entered the horizon sufficiently before matter-radiation equality to reach their asymptotically damped form.

CMBFAST may be modified to incorporate the evolution equation (A.1) and the resulting power spectra calculated. Fig. A.2 shows the effects for the T and B power spectra. The spectra are essentially unchanged at low l . Only at high l do we see a suppression of power from the additional damping. Damping only occurs in waves that have evolved significantly during the radiation-dominated epoch, so this makes sense.

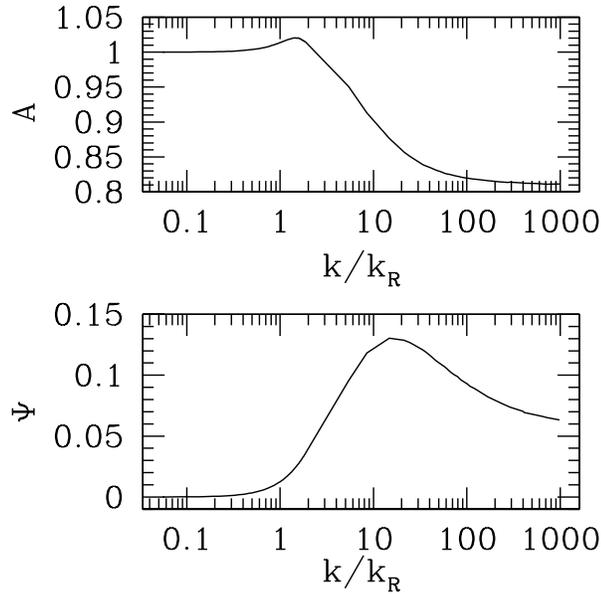


Figure A.1: Transfer function for gravitational waves when anisotropic stress is present. Top panel shows the variation of the amplitude ratio A between the case with anisotropic stress and without anisotropic stress. Bottom panel shows how the phase shift Ψ between the two cases varies with k .

A.2 WKB solution

If the scale factor changes more slowly than the evolution of the gravitational wave, then WKB techniques become a sensible method of approximation. In this Appendix, we detail the application of this approach to the evolution of a gravitational-wave mode through the matter-radiation transition. The WKB approach was first applied in this context by Ng and Speliotopoulos (Ng & Speliotopoulos, 1995), although the solution presented here is of a slightly more general nature.

We begin with the equation of motion for $h(\tau)$ and specific initial conditions that we wish to solve. Working with the dimensionless variables $\eta = (\sqrt{2}-1)\tau/\tau_{\text{eq}}$ and $q = k\tau_{\text{eq}}/(\sqrt{2}-1)$, we have

$$\ddot{h} + 2\frac{\dot{a}}{a}\dot{h} + q^2h = 0, \quad (\text{A.11})$$

$$h(0) = 1, \quad (\text{A.12})$$

$$\dot{h}(0) = 0, \quad (\text{A.13})$$

with overdots indicating differentiation with respect to η . The behaviour of the scale factor

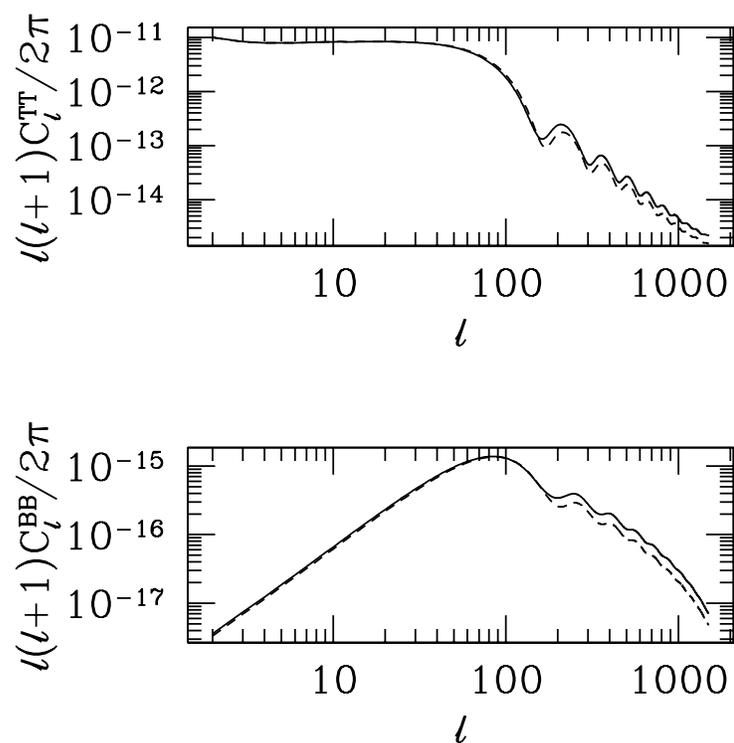


Figure A.2: T and B power spectra incorporating the anisotropic stress term from Eq. (A.1) (dashed curve) and without anisotropic stress (solid curve). Damping in the power is clearly seen on small scales $l > 200$.

in a universe containing only dust and radiation is given by

$$a(\eta) = a_{\text{eq}}\eta(\eta + 2), \quad (\text{A.14})$$

where a_{eq} is the scale factor at equality.

To move towards the standard WKB form, we make the transformation $h = y/[\eta(\eta + 2)]$ which eliminates the first derivative term leading to

$$\ddot{y} + \left(q^2 - \frac{\ddot{a}}{a} \right) = 0. \quad (\text{A.15})$$

Before proceeding, we notice that this transformation is singular at $\eta = 0$, the point at which our boundary conditions are specified. This prevents us from applying the boundary conditions (b.c.) directly and will motivate looking for asymptotic solutions to (A.11) which we will discuss later.

We next map from the interval $[0, \infty]$ to $[-\infty, \infty]$ by the transformations $\eta = \exp(x)$ and $u(x) = \exp(-x/2)y(x)$. These place our problem in the WKB form

$$u''(x) = f(x)u(x), \quad (\text{A.16})$$

$$f(x) = \frac{1}{4} + \frac{2e^x}{e^x + 2} - q^2 e^{2x}. \quad (\text{A.17})$$

The standard WKB problem (Bender & Orszag, 1978)

$$\epsilon^2 y''(x) = Q(x)y(x), \quad (\text{A.18})$$

where $Q(0) = 0$, $Q(x) > 0$ for $x < 0$ and $Q(x) < 0$ for $x > 0$, has the uniform Langer solution

$$y(x) = \sqrt{\pi} \left(\frac{3}{2\epsilon} S_0 \right)^{1/6} [Q(x)]^{-1/4} \left\{ 2C_2 \text{Ai} \left[\left(\frac{3}{2\epsilon} S_0 \right)^{2/3} \right] + C_1 \text{Bi} \left[\left(\frac{3}{2\epsilon} S_0 \right)^{2/3} \right] \right\}, \quad (\text{A.19})$$

with

$$S_0 = \int_x^0 \sqrt{Q(t)} dt. \quad (\text{A.20})$$

Here, $\text{Ai}(x)$ and $\text{Bi}(x)$ are Airy functions and C_1 and C_2 are constant coefficients to be set by the boundary conditions.

Applying this directly to Eqs. (A.16) and (A.17) and manipulating the algebra slightly, we obtain

$$h_{\text{wkb}}(\tau) = \frac{\sqrt{\pi} \Gamma(k\tau)^{-1/4}}{\tau^{1/2}(\tau+2)} \left(\frac{3}{2} S_0(\tau) \right)^{1/6} \times \left\{ 2C_2 \text{Ai} \left[\left(\frac{3}{2} S_0(\tau) \right)^{2/3} \right] + C_1 \text{Bi} \left[\left(\frac{3}{2} S_0(\tau) \right)^{2/3} \right] \right\}, \quad (\text{A.21})$$

with

$$\Gamma(s) = \frac{1}{4} + \frac{2s}{s+2k} - s^2, \quad (\text{A.22})$$

and

$$S_0(\tau) = \int_{k\tau}^{k\tau_T} \sqrt{\Gamma(s)} \frac{ds}{s}. \quad (\text{A.23})$$

Here, τ_T is the solution to $\Gamma(k\tau) = 0$. It can be shown that $k\tau_T$ is bounded such that $1/2 \leq k\tau_T \leq 3/2$. These expressions may be evaluated directly when $\tau < \tau_T$, but some care must be taken when $\tau > \tau_T$. In this case, $\Gamma(k\tau) < 0$, and we must keep careful track of minus signs. Making the appropriate manipulations, we obtain for $\tau > \tau_T$

$$h_{\text{wkb}}(\tau) = \frac{\sqrt{\pi} [-\Gamma(k\tau)]^{-1/4}}{\tau^{1/2}(\tau+2)} \left(\frac{3}{2} S_0(\tau) \right)^{1/6} \times \left\{ 2C_2 \text{Ai} \left[- \left(\frac{3}{2} S_0(\tau) \right)^{2/3} \right] + C_1 \text{Bi} \left[- \left(\frac{3}{2} S_0(\tau) \right)^{2/3} \right] \right\}, \quad (\text{A.24})$$

with

$$\Gamma(s) = \frac{1}{4} + \frac{2s}{s+2k} - s^2, \quad (\text{A.25})$$

and

$$S_0(\tau) = \int_{k\tau_T}^{k\tau} \sqrt{-\Gamma(s)} \frac{ds}{s}. \quad (\text{A.26})$$

Asymptotic expansions of the Airy functions recover the more familiar exponential and trigonometric forms for the WKB connection formula.

Having obtained an expression for h , we now need to apply the boundary conditions. Unfortunately, if we naively try to apply the boundary conditions, we discover that $\dot{h}_{\text{wkb}}(\tau = 0)$ is divergent. This is a consequence of the transformation required to place the equation of motion into WKB form, which is singular at $\tau = 0$. To get around this problem, we seek to apply the boundary conditions at some small time when the gravitational wave has had

little chance to evolve and will be well described by a series solution.

Eq. (A.11) has a regular singular point at $\tau = 0$, so we attempt a Frobenius series solution of the form

$$h(\tau) = \tau^\sigma \sum_{n=0}^{\infty} a_n \tau^n. \quad (\text{A.27})$$

Substituting this expression into Eq. (A.11), solving the indicial equation for σ , and equating like powers of τ to get a recurrence relation leads to an asymptotic polynomial expression for h . In fact, this procedure only generates one of the two linearly independent solutions to Eq. (A.11). Application of the boundary conditions causes the other solution to vanish and normalises this one leaving us with a solution,

$$h(\tau) = 1 - \frac{q^2 \tau^2}{6} + \frac{q^2 \tau^3}{36} - \frac{q^2}{240} (3 - 2q^2) \tau^4. \quad (\text{A.28})$$

This is valid only when $\tau \ll 1$. We can use this to extrapolate the b.c. from zero time to some small time and use it to determine the constant coefficients in Eq. (A.21).

The WKB solution (Fig. A.3) accurately reproduces the phase of the oscillation in the regime $\tau > \tau_T$, although it underestimates the amplitude of h in this region by a factor of ~ 0.87 . The approximation is also good for $\tau < \tau_T$, failing only at times comparable to the time at which the b.c. are applied. For the power spectra calculated in this paper, this is taken to be $k\tau_{\text{match}} = 10^{-5}$. Consequently, the power spectrum become unreliable at small multipoles, $l < 10$. On these scales, power is generated by modes that have evolved very little by the time of recombination.

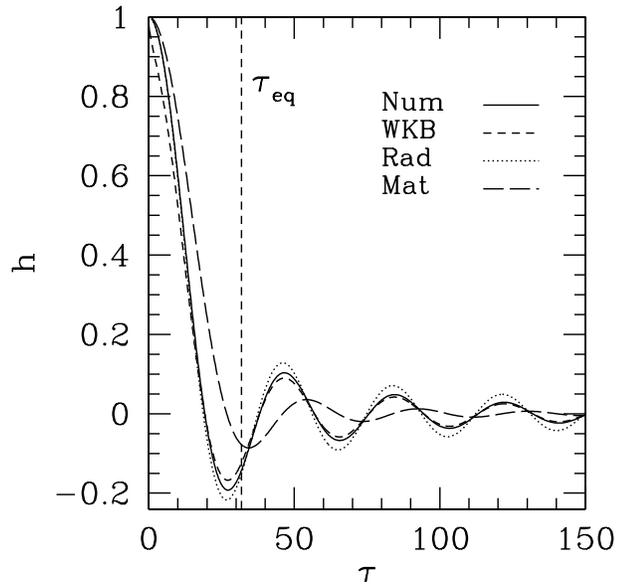


Figure A.3: Comparison of the WKB (Eq. (A.21), dashed curve) and numerical (solid curve) evolution of h for $k = 0.1677 \text{ Mpc}^{-1}$ in the fiducial cosmology. The amplitudes, h_{mat} (Eq. (2.32), long dashed curve) and h_{rad} (Eq. (2.31), dotted curve), are plotted for reference, and τ_{eq} is indicated by a vertical line.

Appendix B

Calculating Einstein A coefficients in the hydrogen atom

The simplicity of the hydrogen atom permits us to compute matrix elements for radiative transitions analytically. We are interested in the Einstein A coefficients, which may be written as (Sobelman, 1972)

$$A(n, l, j, n', l', j') = \frac{64\pi^2}{3h\lambda^3} e^2 (2j' + 1) \left\{ \begin{array}{ccc} l & j & 1/2 \\ j' & l' & 1 \end{array} \right\}^2 l_{>} (R_{n,l}^{n',l'})^2. \quad (\text{B.1})$$

Here $\{...\}$ denotes the Wigner $6 - j$ symbol and we assume spin-1/2 particles. In this expression, the matrix element $R_{n,l}^{n',l'}$ has the usual quantum numbers n , l , and n_r with $n - l - 1 \equiv n_r$. We use the sets (n, n_r, l) and (n', n'_r, l') to describe the upper and lower levels in the spontaneous transition. In addition, we let $l_{>}$ be the greater of l and l' . Because they are separated by $\Delta l = l' - l = \pm 1$ $2l_{>} = l + l' + 1$.

We next define, for $n' - l_{>} \geq 1$,

$$r = n' - l_{>} - 1 \geq 0. \quad (\text{B.2})$$

We also let

$$u = (n - n') / (n + n'); \quad (\text{B.3})$$

$$v = 1 - 1/u^2 = -4nn' / (n - n')^2; \quad (\text{B.4})$$

$$w = v / (v - 1) = 4nn' / (n + n')^2. \quad (\text{B.5})$$

Thus, we only need $R_{n,l}^{n',l'} \equiv a_0 R$, where a_0 is the Bohr radius. This is given by (Rudnick,

1935)

$$R = 2^{l+l'+4} n^{l'+3} n^{l+3} \frac{(n-n')^{n-n'-1}}{(n+n')^{n+n'+1}} \left[\frac{(n+l)!}{(n'+l)!(n-l-1)!(n'-l'-1)!} \right]^{1/2} P_{\pm}. \quad (\text{B.6})$$

Here P_- and P_+ (the subscripts corresponding to the sign of Δl) are terminating hypergeometric series

$$\begin{aligned} 2n'P_- &= (-1)^r [(n-n')^{2r}(2l_{>}+r)!/(2l_{>}!)] \\ &\quad \times [(n+n')F(-r, -n+l_{>}+1; 2l_{>}+1; v) \\ &\quad - (n-n')F(-r, -n+l_{>}; 2l_{>}+1; v)], \quad (\text{B.7}) \end{aligned}$$

$$\begin{aligned} 2nP_+ &= (-1)^r [(n-n')^{2r}(2l_{>}+r)!/(2l_{>}!)] \\ &\quad \times [(n+n')(n-l_{>})F(-r, -n+l_{>}+1; 2l_{>}+1; v) \\ &\quad - (n-n')(n+l_{>})F(-r, -n+l_{>}; 2l_{>}+1; v)]. \quad (\text{B.8}) \end{aligned}$$

These expressions can be inserted into equations (3.9) and (3.12) to compute P_{if} and $f_{\text{recycle}}(n)$. They are in good agreement with existing experimental measurements¹.

¹See reference data at the National Institute of Standards and Technology, <http://physics.nist.gov/>.