Probes of the Early Universe

by

Max Erik Tegmark

B.A., Stockholm School of Economics, 1989
B.A., Royal Institute of Technology, Stockholm, 1990
M.A., University of California at Berkeley, 1992

A dissertation submitted in satisfaction of the final requirement for the degree of

Doctor of Philosophy

in

Physics

in the
Graduate Division
of the

University of California at Berkeley

Committee in charge:

Professor Joseph Silk, Chair
Professor Bernard Sadoulet
Professor Hyron Spinrad

April 1994

The dissertation of Max Tegmark is approved:

Chair Date

Date

Date

University of California at Berkeley

April 1994

Probes of the Early Universe

by

Max Tegmark

Abstract

Probes of the Early Universe

by

Max Erik Tegmark

Doctor of Philosophy in Physics

University of California at Berkeley

Professor Joseph Silk, Chair

One of the main challenges in cosmology is to quantify how small density fluctuations at the recombination epoch $z \approx 10^3$ evolved into the galaxies and the large-scale structure we observe in the universe today. This thesis discusses ways of probing the intermediate epoch, focusing on the thermal history. The main emphasis is on the role played by non-linear feedback, where a small fraction of matter forming luminous objects can inject enough energy into the inter-galactic medium (IGM) to radically alter subsequent events. The main conclusions are:

- Early structures corresponding to rare Gaussian peaks in a cold dark matter (CDM) model can photoionize the IGM early enough to appreciably smooth out fluctuations in the cosmic microwave background radiation (CBR), provided that these early structures are quite small, no more massive than about $10^8 M_{\odot}$.

- Typical parameter values predict that reionization occurs around $z = 50$, thereby reducing fluctuations on degree scales while leaving the larger angular scales probed by COBE relatively unaffected.

- For non-standard CDM incorporating mixed dark matter, vacuum density, a tilted primordial power spectrum or decaying τ neutrinos, early reionization is not likely to play a significant role.

- For CDM models with $\Omega < 1$ and $\Lambda = 0$, the extent of this suppression is quite insensitive to Ω_0, as opposing effects partially cancel.

- It is still not certain that the universe underwent a neutral phase, despite the new COBE FIRAS limit $y < 2.5 \times 10^{-5}$ on Compton y-distortions of the cosmic microwave background.

- The observed absence of a Gunn-Peterson trough in the spectra of high-redshift quasars can be explained without photoionization, in in a scenario in which supernova-driven winds from early galaxies reionize the IGM by $z = 5$.

- It is possible to place constraints on cosmological models that are independent of the shape of the primordial power spectrum — the only assumption being that the random fields are Gaussian. As an example of an application, the recent measurement of bulk flows of galaxies by Lauer and Postman is shown to be inconsistent with the CBR experiment SP91 at a 95% confidence level regardless of the shape of the power spectrum.

Chair Date

To Richard Feynman,

1918-1988,

who was the reason I decided to switch to physics

Contents

List of Figures

List of Tables

List of Abbreviations

BBKS Bardeen, Bond, Kaiser & Szalay (1986), a reference in Chapter 6.

BDM Baryonic Dark Matter

CBR Cosmic (Microwave) Background Radiation

CDM Cold Dark Matter

COBE Cosmic Background Explorer

DMR Differential Microwave Radiometer, an instrument on board the COBE satellite that measures the difference in CBR temperature between different parts of the sky

FIRAS Far Infra-Red Absolute Spectrometer, an instrument on board the COBE satellite that measures the CBR spectrum

FRW The Friedmann-Robertson-Walker metric, sometimes referred to as the Friedmann-Lemaître metric

GO Gnedin & Ostriker (1993), a reference in Chapter 5

GR General Relativity

HDM Hot Dark Matter

IGM Intergalactic Medium

IRAS Infra-Red Astronomical Satellite

ΛCDM Cold Dark Matter with a non-zero cosmological constant Λ

LP Lauer & Postman, a reference in Chapter 7

MACHO Massive Compact Halo Object, a dark matter candidate

MDM Mixed Dark Matter, a mixture of HDM and CDM

SP91 South Pole 1991 CBR experiment referenced in chapters 3, 4 and 7

QSO Quasi-Stellar Object — quasars are a subset of these

τCDM Cold Dark Matter with a decaying τ neutrino

UV Ultraviolet

WIMP Weakly Interacting Massive Particle, a dark matter candidate

Preface

As this thesis is laden with cosmology and astrophysics jargon, it is recommended that the non-cosmologist reader flick through Chapter 2 before attempting to read anything else. This chapter is by no means intended to be a comprehensive treatise on the foundations of cosmology. It is merely an attempt to provide the basic knowledge necessary for a reader with a general physics background to be able to follow the other chapters. Cosmologists are advised to skip directly from the introduction to Chapter 3.

During my graduate studies here at Berkeley, I also did some non-cosmology work (Tegmark 1993, Tegmark & Yeh 1994, Tegmark & Shapiro 1994, Shapiro & Tegmark 1994), but that research will not be discussed in this thesis.

Acknowledgements...

I wish to thank

my advisor Joe Silk

> for the inspiration, support and guidance,
> for the infinite degrees of freedom, and
> for proving that success in science is not irreconcilable with having a life,

my Mom,

> one foxy lady,
> thanks to whom science is still less than half of my life,

my Dad,

> who encouraged my curiosity and shared my fascination
> and who thinks that the universe was created as a source of interesting math problems,

my brother Per

> — iiiiiiih...,

Justin Bendich, Ted Bunn, Wayne Hu, Bill Poirier and Leehwa Yeh,

> with whom I've so enjoyed sharing my bizarre physics interests, be it over French Alps or homebrewn beer,

teachers like Geoffrey Chew, Andy Elby, Alex Filippenko,
Lawrence Hall and Bruno Zumino, who inspired me,

the Cyberscope and the Simsalabimiers, who stole away so much of my time,

the friendly staff in the physics and astronomy departments here at Berkeley,
especially Anne Takizawa and Donna Sakima,
who saved me every time I was dropped from the rolls of the university,

and last but not least, the people close to me here in Berkeley,
who have put up with this crazy cosmologist and made these four years
perhaps the happiest of my life.

...and More Acknowledgements

The research presented in Chapters 3 through 7 was done in collaborations, all but Chapter 7 with my thesis advisor Joe Silk, Chapter 3 with Alain Blanchard, Chapter 6 with August Evrard, and Chapter 7 with Emory Bunn and Wayne Hu. In all cases except Chapter 7, I made the calculations, produced the figures and wrote the bulk of the text. In "Power spectrum independent constraints on cosmological models", Emory Bunn performed the Monte Carlo analysis for modeling the COBE experiment and made the fits described by equations (7.2) and (7.3). He also calculated all the window functions analytically and numerically, wrote Appendix B and wrote the first version of our FORTRAN data analysis code. Wayne Hu originally spawned the idea of applying power spectrum independent constraints to bulk flows, and independently verified the calculations.

For Chapter 2, I gratefully acknowledge Naoshi Sugiyama for letting me use power spectra generated by his Boltzmann code in Figure 2.3 and Martin White for letting me use window function data from White, Scott & Silk (1994) in Figure 2.5.

For Chapters 3 and 5, I gratefully acknowledge William Vacca for providing stellar spectra used to calculate an entry in Table 3.3 and Table 5.1.

In addition, I wish to thank Emory Bunn, Angelica de Oliveira Costa, Wayne Hu, Andreas Reisenegger, David Schlegel, Douglas Scott, Harold Shapiro, Joseph Silk, Charles Steidel, Michael Strauss, and Martin White, for many useful comments and/or for help with proofreading the manuscript.

Chapter 1

Introduction

One of the main challenges in modern cosmology is to quantify how small density fluctuations at the recombination epoch, about half a million years after the Big Bang, evolved into the galaxies and the large-scale structure we observe in the universe today, some 10^{10} years later. This thesis focuses on ways of probing the interesting intermediate epoch. The main emphasis is on the role played by non-linear feedback, where a small fraction of matter forming luminous objects such as stars or QSO's can inject enough energy into their surroundings to radically alter subsequent events. Specific questions addressed include the following:

- Which cosmological models predict that the universe will become photoionized early enough to measurably affect spatial fluctuations in the cosmic microwave background radiation (CBR)?
- Which cosmological models are ruled out by the constraints from the COBE FIRAS experiment on the spectral distortions of the CBR?
- Which cosmological models satisfy the Gunn-Peterson test, *i.e.* predict the intergalactic medium (IGM) to become highly ionized by $z = 5$?
- Is reionization required to make bulk flows consistent with degree-scale anisotropies?

Thus this thesis provides a quantitative link between cosmological models and extragalactic observational astronomy, showing how current observations constrain the allowed parameter space of cosmological models.

The next two sections of this introduction are aimed at placing the work in a larger context, and the final section gives a brief description of how the thesis is organized into chapters. As mentioned in the preface, the non-cosmologist is advised to flick through Chapter 2 first, before attempting to read further.

1.1 Cosmology towards a New Millennium

Like the universe, the field of cosmology is expanding. Thanks to the rapid technological developments of the last few decades, we humans now have better eyes than ever before with which to scrutinize our surroundings, both from the ground and from various spacecraft.

On the ground, the recently inaugurated ten-meter Keck telescope on Hawaii provides record-breaking light-gathering power in optical wavelengths. Interferometry between radio telescopes on separate continents provides angular resolutions that our ancestors could only have dreamed of. Semiconductor technology has replaced photographic plates by much more sensitive electronic detectors (CCDs), and also, in the form of computers, enabled previously unheard of capacity for data processing. A notable example is the MACHO project, looking for dark matter gravitational microlensing events in the halo of our galaxy, a search that involves looking at some one million stars every night (Alcock *et al.* 1993). Another example is the Sloan Digital Sky Survey, whose goal is to measure the redshifts of a million galaxies before the end of the millennium. To place this in perspective, when Slipher obtained the the 18 data points that Hubble used to support his famous expansion law (Hubble 1929), he required weeks of observations with a by contemporary standards good telescope in order to obtain the redshift of a single nearby galaxy.

In space, scores of detectors have been observing the cosmos unhindered by the astronomer's worst enemy: the earth's atmosphere. After being repaired, the Hubble Space Telescope now obtains optical images with ten times better angular resolution than obtainable from the ground. The Cosmic Background Explorer satellite (COBE) has made an all-sky map at millimeter wavelengths (Smoot *et al.* 1992). The IRAS and ROSAT satellites have done the same in the infrared and x-ray wavelength bands, respectively, and a number of satellites have explored gamma rays and the ultraviolet as well. Many of these wavelengths are almost completely absorbed by the atmosphere, and thus inaccessible to ground-based experiments. An overview of our knowledge of the cosmological photon spectrum in all wavelength bands from 10^{-24}cm to 10^{6}cm is given by Ressell & Turner (1990).

With all this data pouring in, it is a daunting task for theorists to keep up and try to make sense of it all. Although it cannot be overemphasized how much remains to be done, the advances (interrupted by occasional false clues) of the last few decades have been quite astonishing by any standards. As a more detailed example, let us take something that is directly relevant to the material in this thesis: the cosmic microwave background radiation (CBR). Thirty years ago, we did not even know that it existed. Confirming theoretical predictions of Gamow and others, it was serendipitously discovered by Penzias and Wilson in 1964 (Penzias & Wilson 1965) and given its cosmological interpretation by Dicke, Peebles, Roll & Wilkinson (1965). In Figure 1.1, the best limit on the Compton y-parameter is plotted as a function of year. This parameter is a measure of how much the spectrum deviates from a Planck spectrum, and when it was first defined (Zel'dovich & Sunyaev 1969), one was still arguing as to whether the CBR had a Planck spectrum or not, so the constraints on y were of order unity. About a decade later, the limit had dropped to about 0.25 (Woody & Richards 1981). Almost another decade later, it dropped to around 0.1 thanks to a Berkeley-Nagoya rocket experiment (Matsumoto *et al.* 1988). Here a fairly common cosmo-sociological effect occurred: an apparent deviation from a Planck spectrum, now sarcastically referred to as "the Berkeley Bump", led to scores of theoretical models explaining it before it was officially admitted to be merely an artifact caused by problems with the detector. In 1990, the FIRAS experiment aboard the COBE satellite suddenly pushed the limit down by a full two orders of magnitude (Mather *et al.* 1990). And the exponential progress continued, the current limit being 2.5×10^{-5} (Mather *et al.* 1994).

This mind-boggling progress with CBR spectral distortions has been paralleled by improved measurements of the spatial distortions. The fairly isotropic radiation field was observed to have a dipole anisotropy of order 10^{-3} (Conklin 1969; Henry 1971; Smoot *et al.* 1977), which is normally interpreted as a Doppler shift due to our motion relative to the CBR rest frame. A few years ago, the COBE DMR experiment detected smaller scale fluctuations as well (Smoot *et al.* 1992), at a level of about 10^{-5}, which are believed to be of cosmological origin. Since this time period, there has been an enormous surge in balloon-based and ground-based CBR experiments that probe smaller angular scales, and now in the spring of 1994, new results are coming in at a rate of almost one per month.

It is precisely this type of data, in combination with improved constraints on the matter power spectrum from galaxy surveys, that is allowing us to start pinning down the thermal history of the IGM since recombination, the main topic of this thesis.

1.2 Reheating

It is now widely accepted that the universe underwent a reheating phase at some point after the standard recombination epoch at redshift $z \approx 10^3$. The absence of a Gunn-Peterson trough in the spectra of high-redshift quasars has provided strong evidence for the reheating occurring at a redshift $z > 5$, since it indicates that the intergalactic medium (IGM) was highly ionized at lower redshifts (Gunn & Peterson 1965; Steidel & Sargent 1987; Webb *et al.* 1992). The smallest baryonic objects to go non-linear in a standard cold dark matter (CDM) model are expected to reionize the IGM at a redshift somewhere in the range $20 < z < 100$ (Couchman & Rees 1986). In the most recent models with baryonic dark matter, reheating and reionization are predicted to occur at an even higher redshift, typically in the range $100 < z < 1000$ (Peebles 1987; Gnedin & Ostriker 1992; Cen *et al.* 1992; Cen *et al.* 1993).

A reheating epoch would have at least two interesting classes of effects that may be measurable today: effects on subsequent structure formation and effects on the cosmic microwave background radiation (CBR).

Subsequent structure formation would be affected in a number of ways. First of all, the heating of the IGM up to a higher adiabat would raise the Jeans mass, thus suppressing the formation of small objects. For instance, an IGM temperature of 10^5K at a redshift of a few would suppress the formation of galaxies of mass below $10^{10} M_\odot$, thus alleviating the ubiquitous problem of theories overpredicting the abundance of faint galaxies (Blanchard *et al.* 1992). Secondly, if the objects that reheat the IGM also enrich it with heavy elements, the ability of gas to cool would be greatly enhanced in the temperature range $10^4\text{K} < T < 10^7\text{K}$, presumably facilitating future structure formation.

The CBR would be affected in at least three ways:

1. Hot ionized IGM would cause spectral distortions (Kompaneets 1957; Zel'dovich &

Sunyaev 1969; Bartlett & Stebbins 1991) which might violate the stringent limits on the the Compton y-parameter (Mather *et al.* 1994).

2. Spatial fluctuations on angular scales below a few degrees might be suppressed, while fluctuations on larger scales would remain fairly unaffected. This effect is particularly interesting since some degree-scale microwave background experiments carried out at the South Pole (Meinhold & Lubin 1991; Shuster *et al.* 1993) and at balloon altitudes (Devlin *et al.* 1992; Meinhold *et al.* 1993; Shuster *et al.* 1993) appear to detect fluctuation amplitudes lower than those expected in a CDM model normalized to the the COBE DMR experiment (Smoot *et al.* 1992), which probes larger angular scales. The interpretation of this remains unclear, however, as many other experiments (*e.g.* de Bernardis *et al.* 1993; Dragovan *et al.* 1993; Cheng *et al.* 1993; Gundersen *et al.* 1993) have detected higher levels of fluctuations.

3. New spatial fluctuations will be generated on smaller angular scales, through the so called Vishniac effect (Vishniac 1987, Hu, Scott & Silk 1994). The current upper limit on CBR fluctuations on arcminute scales (*e.g.* Subrahmanyan *et al.* 1993) places constraints on some reheating scenarios.

With the recent surge in CBR experiments and the many new numerical, theoretical and observational results on structure formation, the thermal history of the universe is now coming within better reach of our experimental probes. In view of this, it is very timely to theoretically investigate the nature of the reheating epoch in greater detail. It is also important to use observational data to place firm constraints on the thermal history of the universe, constraints that do not require any assumptions about uncertain entities like the primordial power spectrum.

1.3 A Sneak Preview

In this thesis, we will do the following:

In Chapter 3, early photoionization of the intergalactic medium is discussed in a nearly model-independent way, in order to investigate whether early structures corresponding to rare Gaussian peaks in a CDM model can photoionize the intergalactic medium early enough to appreciably smooth out the microwave background fluctuations. We conclude that this is indeed possible for a broad range of CDM normalizations and is almost inevitable for unbiased CDM, provided that the bulk of these early structures are quite small, no more massive than about $10^8 M_\odot$. Typical parameter values predict that reionization occurs around $z = 50$, thereby suppressing fluctuations on degree scales while leaving the larger angular scales probed by COBE relatively unaffected. However, for non-standard CDM, incorporating mixed dark matter, vacuum density or a tilted primordial power spectrum, early reionization plays no significant role.

In Chapter 4, the results of Chapter 3 are generalized to CDM models where $\Omega < 1$ and the cosmological constant $\Lambda > 0$. Such models often require early reionization to suppress degree-scale anisotropies in order to be consistent with experimental data. It is found that if the cosmological constant $\Lambda = 0$, the extent of this suppression is quite insensitive to Ω_0, as opposing effects partially cancel. Given a σ_8-normalization today, the loss of small-scale power associated with a lower Ω_0 is partially canceled by higher optical depth from longer lookback times and by structures forming at higher redshifts as the universe becomes curvature dominated at $z \approx \Omega_0^{-1}$. The maximum angular scale on which fluctuations are suppressed decreases when Ω_0 is lowered, but this effect is also rather weak and unlikely to be measurable in the near future. For flat models, on the other hand, where $\lambda_0 = 1 - \Omega_0$, the negative effects of lowering Ω_0 dominate, and early reionization is not likely to play a significant role if $\Omega_0 \ll 1$. The same goes for CDM models where the shape parameter Γ is lowered by increasing the number of relativistic particle species.

In Chapter 5, we find that it is still not certain that the universe underwent a neutral phase, despite the new COBE FIRAS limit $y < 2.5 \times 10^{-5}$ on Compton y-distortions of the cosmic microwave background. Although scenarios where the very early ($z \sim 1,000$) ionization is thermal (caused by IGM temperatures exceeding 10^4K) are clearly ruled out, there is a significant loophole for cosmologies with typical CDM parameters if the dominant ionization mechanism is photoionization. If the ionizing radiation has a typical quasar spectrum, then the y-constraint implies roughly $h^{4/3}\Omega_b\Omega_0^{-0.28} < 0.06$ for fully ionized models. This means that baryonic dark matter (BDM) models with $\Omega_0 \approx 0.15$ and reionization at $z \approx 1,000$ are strongly constrained even in this very conservative case, and can survive the y test only if most of the baryons form BDM around the reionization epoch.

In Chapter 6, a model is presented in which supernova-driven winds from early galaxies reionize the IGM by $z = 5$. This scenario can explain the observed absence of a Gunn-Peterson trough in the spectra of high-redshift quasars providing that the bulk of these early galaxies are quite small, no more massive than about $10^8 M_\odot$. It also predicts that most of the IGM was enriched to at least 10% of current metal content by $z = 5$ and perhaps as early as $z = 15$. The existence of such early mini-galaxies violates no spectral constraints and is consistent with a pure CDM model with $b \leq 2$. Since the final radius of a typical ionized bubble is only around 100 kpc, the induced modification of the galaxy autocorrelation function is negligible, as is the induced angular smoothing of the CBR. Some of the gas swept up by shells may be observable as pressure-supported Lyman-alpha forest clouds.

In Chapter 7, a formalism is presented that allows cosmological experiments to be tested for consistency, and allows a simple frequentist interpretation of the resulting significance levels. As an example of an application, this formalism is used to place constraints on bulk flows of galaxies using the results of the microwave background anisotropy experiments COBE and SP91, and a few simplifying approximations about the experimental window functions. It is found that if taken at face value, with the quoted errors, the recent detection by Lauer and Postman (1994) of a bulk flow of 689 km/s on scales of $150h^{-1}$Mpc is inconsistent with SP91 at a 95% confidence level within the framework of a Cold Dark

Matter (CDM) model. The same consistency test is also used to place constraints that are completely model-independent, in the sense that they hold for any power spectrum whatsoever — the only assumption being that the random fields are Gaussian. It is shown that the resulting infinite-dimensional optimization problem reduces to a set of coupled non-linear equations that can readily be solved numerically. Applying this technique to the above-mentioned example, we find that the Lauer and Postman result is inconsistent with SP91 even if no assumptions whatsoever are made about the power spectrum.

These chapters are all self-contained and fairly independent (with the exception of 3 and 4), so they can be read in any order.

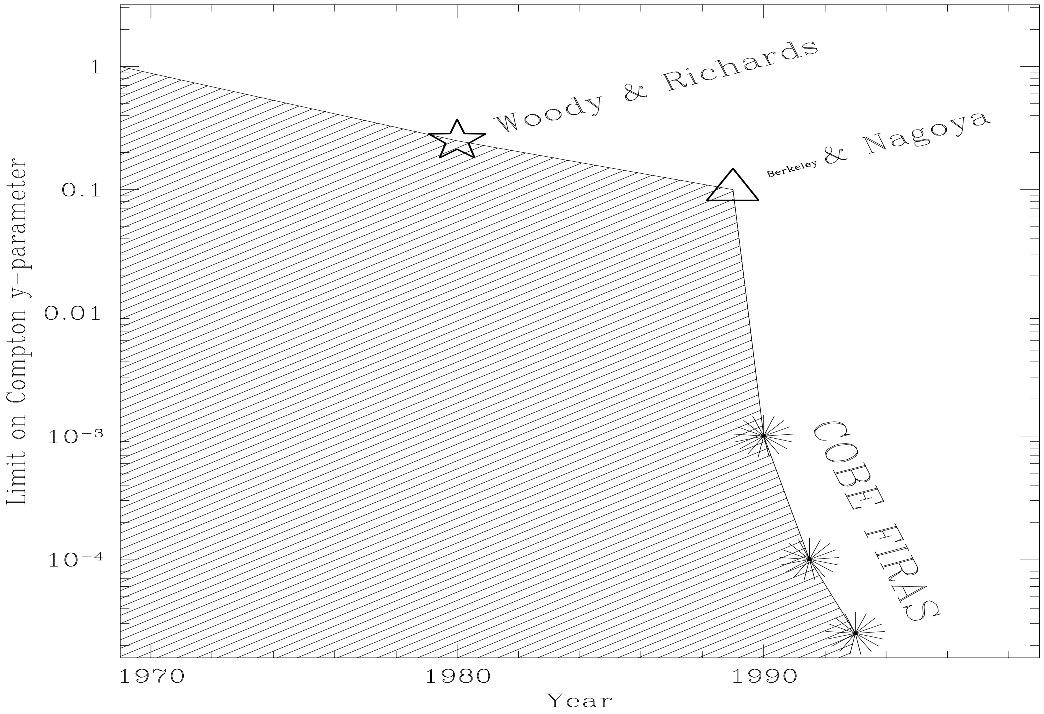


Figure 1.1: Limits on the Compton y-parameter.

The observational upper limit on the Compton y-parameter is plotted as a function of year.

Chapter 2

A Cosmology Primer

This chapter is by no means intended to be a thorough and systematic treatise on the basics of cosmology. It is merely an attempt to provide the bare-bones knowledge necessary for the reader with a general physics background to be able to follow the other chapters. Cosmologists are advised to skip directly to Chapter 3.

2.1 Where to Read More

For the reader interested in the basics of modern cosmology, there are many excellent books to chose from. Two early classics are

- *Physical Cosmology* by Peebles (1971) and
- *Gravitation and Cosmology* by Weinberg (1972).

More recent texts include

- *The Early Universe* by Kolb and Turner (1990),
- *Modern Cosmology* by Dolgov, Sazhin and Zel'dovich (1990),
- *Introduction to Cosmology* by Narlikar (1993)
- *Structure Formation in the Universe* by Padmanabhan (1993) and
- *Principles of Physical Cosmology* by Peebles (1993).

There is also a large number of good review articles, for instance in conference proceedings and printed lecture series. A very incomplete list includes the following:

- *Vatican Study Week: Large Scale Motions in the Universe*, edited by Rubin & Coyne (1988),
- *Development of Large-Scale Structure in the Universe* by Ostriker (1991),
- *Physics of the Early Universe*, edited by Peacock, Heavens and Davies 1990 (see especially the articles by Efstathiou and White),

- *Observational and Physical Cosmology*, edited by Sánchez, Collados and Rebolo 1990 (see especially the article by Bernard Jones),

The detailed reference information for these publications is in the reference section at the end. Essentially all the results in the rest of this chapter can be found in books such as these, so very few explicit references will be given.

2.2 So what is cosmology all about?

Cosmologists study the history of the universe, from its birth some 10^{10} years ago up until today. Since the discovery of the cosmic microwave background radiation, the Big Bang model has been generally accepted by the scientific community. Although we still know very little about what happed before the universe was about a hundredth of a second old, the Big Bang model is widely believed to give a correct basic picture of what happened after that. Here are some of the cornerstones of this standard cosmological model:

1. The universe is homogeneous and isotropic on large scales.
2. The universe is expanding.
3. The universe used to be dense and hot.
4. The universe is full of microwaves.
5. Structures formed through gravitational instability.
6. The universe contains dark matter.

We will now discuss these issues one by one.

1. By large-scale homogeneity is meant that if we average over sufficiently large volumes, the density of the universe is roughly the same everywhere. This of course implies that the density is isotropic as well, *i.e.* that it appears the same in whatever direction we choose to look[1]. Not only the density, but all other fields as well (the velocity field, the metric, *etc.*) are assumed to be homogeneous and isotropic on large scales. There is now strong observational support for 1. Apart from this, we also tend to believe in homogeneity and isotropy because this greatly simplifies our calculations...

2. All distant galaxies that we see in the sky are moving away from us, and the further away they are, the faster they tend to be receding. The homogeneity implies that on large scales, everything is indeed receding from everything else. In terms of classical physics, this is analogous to the way that all the raisins in an expanding raisin cake recede from one another, and just as in this analogy, the speed with which to objects are receding from each other is proportional to the distance between them. The proportionality constant H is called the Hubble constant. Its value is such that objects separated by more than about 10^{10}

[1]Note that homogeneity implies isotropy only for scalar quantities (such as the density), not vectorial or tensorial quantities such as the velocity or the metric.

light-years are receding from each other faster than the speed of light. Doesn't this violate special relativity? Yes it does, but it's OK in general relativity... As will be discussed in the next section, rather than picturing the matter expanding through space at enormous speeds, a more natural way to think of this is to imagine the the matter as merely sitting still while space itself is expanding. The mathematical description of the expanding universe is called the Friedmann-Robertson-Walker (FRW) metric, and will be introduced in the next section.

3. Extrapolating this expansion backwards in time, one concludes that the universe used to be much denser and hotter in the past. Indeed, the FRW metric tells us that some 10^{10} years ago[2], the universe was denser and hotter than any object we have ever been able to study. Since we have no experimental knowledge about how matter behaves when it is so hot and dense, there was an early phase of the universe about which we cannot predict anything without making speculative extrapolations. After the universe was about a hundredth of a second old, however, the temperatures were no greater than those that we can study in stars and in laboratories, so the physical processes that operated after that are fairly well understood and can be used to make detailed quantitative predictions. One such prediction is that most of the mass of the universe will be in the form of hydrogen, and about 25% will be helium. The fact that the observed helium fraction is indeed around 25% is one of the great triumphs of the Big Bang model.

4. Another such prediction is that the universe is full of microwaves. These microwaves have roughly the same wavelength as those that heat our breakfast hot dogs, but their density is much lower and their spectral distribution is almost exactly that which would be emitted by a black object whose temperature was a few degrees Kelvin, *i.e.* just a few degrees above absolute zero. The discovery of this cosmic microwave background radiation in 1964 was another great triumph of the Big Bang model, arguably the greatest of them all.

5. By studying these microwaves, we can infer that the density of the universe was almost perfectly uniform about half a million years after the Bang. Today, it is full of dense clumps, ranging in size from superclusters of galaxies to the planet we live on, not to mention the author. The main mechanism responsible for turning the primordial ordered state into such a mess is believed to be gravitational instability, whereby the gravity from regions slightly denser than average attracted surrounding matter, gradually creating ever larger lumps. This phenomenon is usually modeled as perturbations to the FRW metric, and will be discussed in Section 2.4.

6. By studying the rotation of different parts of our galaxy, we can calculate how much mass there is at various distances from the galactic center. The surprising conclusion is

[2] An amusing fact is that when Edwin Hubble first attempted to measure the constant that now bears his name, a false assumption led him to calculate a value more than five times larger than the current estimates. This implied that the universe was "only" about 2,000,000,000 years old. Since radioactive dating methods had showed certain meteorites to be about twice as old as that, the Big Bang theory came under fire from people wondering how rocks could be older than the universe itself. This age discrepancy was one of the reasons that Hoyle and others developed a competing (and now dead) theory, the so called steady state theory, in which there is no Big Bang — and indeed the reason that Hoyle condescendingly coined the term "Big Bang".

that there is much more mass present than can be accounted for by the matter we see. The mysterious invisible substance (or substances) is referred to as *dark matter*. There are also other indications of dark matter. Figuring out its nature and its density is one of the hottest problems in modern cosmology (Sadoulet & Cronin 1991), as its gravitational presence affects the formation of structure mentioned in 5.

The only parts of cosmology and astrophysics with which the reader should be familiar to be able to follow the following chapters are

- The standard model of space and time in an expanding universe, including quantities such as the redshift z, the density parameter Ω and the cosmological constant Λ
- The basics of structure formation, including quantities like the random field $\delta(\mathbf{x})$ and the matter power spectrum $P(k)$
- The microwave background, including quantities such as the the radiation power spectrum and window functions
- Rudimentary hydrogen chemistry
- Some basic cosmology jargon, such as units like Mpc and fudge factors like h

These will be the topics of the following five sections.

2.3 Bare Bones General Relativity and the FRW Metric

In this section, we will say a few words about general relativity and the large-scale structure of space-time, as well as list a number of handy formulas that will be used in subsequent chapters.

2.3.1 General Relativity

> Journalist: *"Professor Eddington, is it really true that only three people in the world understand Einstein's theory of general relativity?"*
> Eddington: *"Who is the third?"*

It would obviously be absurd to try to squeeze a real presentation of general relativity, one of the deepest physical theories ever to be invented by man, into a subsection of a Ph.D. thesis. In fact, I have been asked how so many cosmologists can use a theory as mathematically complex as "GR" in their work without even knowing basic differential geometry. I think that the answer is that some aspects of GR are quite intuitive, so that one can acquire a working feeling for what is going on even without being able to tell a Lie derivative from a Killing field. In addition to this, the FRW metric is so much simpler than the most generic solutions that what happens on distance scales much smaller than 10^{10} light-years can often be adequately described by classical physics plus some geometric considerations. In fact, in some ways GR resembles classical physics more than it resembles special relativity.

2.3.2 The raw equations

So how do you calculate things with GR? What is the basic mathematical structure of the theory? The basic objects are some fields that live on a four-dimensional manifold, and satisfy a bunch of messy nonlinear second order partial differential equations. The *Einstein field equations* are[3]

$$R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} = 8\pi G T_{\mu\nu},$$

where

$$R \equiv g^{\mu\nu} R_{\mu\nu},$$

$$R_{\mu\nu} \equiv R^{\alpha}_{\mu\alpha\nu},$$

$$R^{\alpha}_{\mu\nu\beta} = \Gamma^{\alpha}_{\nu\beta,\mu} - \Gamma^{\alpha}_{\mu\beta,\nu} + \Gamma^{\gamma}_{\mu\beta}\Gamma^{\alpha}_{\nu\gamma} - \Gamma^{\gamma}_{\nu\beta}\Gamma^{\alpha}_{\mu\gamma},$$

$$\Gamma^{\alpha}_{\mu\nu} = \frac{1}{2} g^{\alpha\sigma}\left(g_{\sigma\mu,\nu} + g_{\sigma\nu,\mu} - g_{\mu\nu,\sigma}\right),$$

and $g^{\mu\nu}$ is the matrix inverse of $g_{\mu\nu}$, *i.e.*

$$g^{\mu\alpha} g_{\alpha\nu} = \delta^{\mu}_{\nu}.$$

[3] A number of different notational conventions can be found in the literature. Here we use the sign convention $(+ - - -)$ for the metric. In some texts, the signs of the last two terms in the definition of of $R^{\alpha}_{\mu\nu\beta}$ are reversed, which corresponds to a different convention for the index placement.

Here repeated indices are to be summed over from 0 to 3, commas denote derivatives as in the standard tensor notation, and G is the gravitational constant. Throughout this section, we will use units where the speed of light $c = 1$. In the Einstein field equations, the dependent variables are the two tensors $g_{\mu\nu}$ and $T_{\mu\nu}$. They are both symmetric, and thus contain ten independent components each. g is called the *metric tensor*, and describes the structure of spacetime at each spacetime point x^μ. T is called the *stress-energy tensor*, and describes the state of the matter (what is *in* space) at each point. The quantities $\Gamma^\alpha_{\mu\nu}$, $R^\alpha_{\mu\nu\beta}$ and $R_{\mu\nu}$ are named after Christoffel, Riemann and Ricci, respectively, and will not be further discussed here.

2.3.3 The FRW metric

Truth is much too complicated to allow anything but approximations.
John von Neumann

The differential quantity

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu$$

is called the *line element*, and for the Friedmann-Robertson-Walker (FRW) metric takes the simple form

$$ds^2 = dt^2 - a(t)^2 \left[\frac{dr^2}{1-kr^2} + r^2 d\theta^2 + r^2 \sin^2\theta \, d\varphi^2 \right],$$

where k is a constant that is either zero or ± 1, and a is a function to be determined later. For reasons that will soon become clear, we have renamed the independent variables

$$x^\mu = (t, \mathbf{r}),$$

and changed to spherical coordinates

$$\mathbf{r} = (x_1, x_2, x_3) = (\sin\theta\cos\varphi, \sin\theta\sin\varphi, \cos\theta) r.$$

2.3.4 Interpreting the FRW coordinates

The Newtonian equation of motion for a point particle moving along a trajectory $\mathbf{x}(t)$ in a gravitational field $\mathbf{g}$ is

$$\ddot{x}_i = g_i,$$

where $i = 1, 2, 3$. The analog of this in General Relativity for a point particle moving along a trajectory $x^\mu(t)$ in spacetime, t now being any parameter, is the geodesic equation of motion

$$\ddot{x}^\mu = \Gamma^\mu_{\nu\lambda} \dot{x}^\nu \dot{x}^\lambda.$$

For the FRW metric, a cumbersome but straightforward calculation shows that

$$x^\mu = (t, \mathbf{r}_0),$$

is a solution to the geodesic equation for any constant $\mathbf{r}_0$. This means that if we interpret t as time[4] and call the vector $\mathbf{r}$ the *comoving spatial coordinates*, then the FRW metric can be interpreted as follows: an object at a comoving position $\mathbf{r} = \mathbf{r}_0$ that is at rest (defined as $\dot{\mathbf{r}} = 0$) will remain at rest. The time t is simply the time that such an object would measure if it were a clock. By definition, the physical distance between two comoving points $\mathbf{r}_1$ and $\mathbf{r}_2$ at some time t equals

$$\int_1^2 \sqrt{-ds^2},$$

where the integral is to be taken along the *geodesic*, the curve from $(t, \mathbf{r}_1)$ to $(t, \mathbf{r}_2)$ that minimizes this quantity. A quick look at the expression for ds^2 above tells us that this distance will be proportional to $a(t)$. Thus objects that are at rest relative to the comoving coordinates nonetheless move relative to each other, the time-dependence of all distances scaling as the function $a(t)$, which is named the *scale factor.* When our universe is modeled by the FRW metric, distant galaxies are assumed to stay at fixed comoving coordinates, and the universe expanding simply corresponds to $a(t)$ increasing with time. The Hubble constant is evidently given by

$$H \equiv \frac{\dot{a}}{a}.$$

(Note that calling it a "constant" is a misnomer, as it generally changes with time.)

2.3.5 Curvature

Let us for a moment keep t fixed and say a few words about the constant k. If $k = 0$, then the geodesic (shortest curve) between two points $\mathbf{r}_1$ and $\mathbf{r}_2$ in $\mathbf{R}^3$ will be the conventional straight line. This is not the case when $k \neq 0$. If we chose three points $\mathbf{r}_1$, $\mathbf{r}_2$ and $\mathbf{r}_3$ and connect them in a triangle by the three geodesics, the sum of the three angles will be greater that $180°$ if $k = 1$ (in which case space is said to be *positively curved*), equal to $180°$ if $k = 0$ (in which case space is said to be *flat*), and less than $180°$ if $k = -1$ (in which case space is said to be *negatively curved* or *hyperbolic*). If $k \leq 0$, space is infinite at all times[5], the coordinate $\mathbf{r}$ is allowed to range over the entire space $\mathbf{R}^3$, and the universe is said to be *open.* If $k = 1$, however, space has the topology of the 3-sphere, the universe is said to be closed, and its volume is finite for any t. In this case, the comoving spatial coordinates are restricted to the range $|\mathbf{r}| < 1$. (This coordinate patch covers exactly half of the three-sphere.)

2.3.6 Redshift

Let us now turn to the time-dependence. One can show that the momentum of all free particles decreases as the inverse of the scale factor a as the FRW universe expands. Thus massive particles gradually slow down and become comoving, "go with the flow", ending up at rest with respect to the comoving coordinates. For photons, which cannot slow down, the momentum loss corresponds to a lowering of their frequency, *i.e.* they become redshifted.

[4]This is a very fortunate feature of the FRW metric, since for generic metrics, it turns out to be impossible to define a global time that observers in different parts of the universe can agree on.

[5]Except in unusual topologies that will not be discussed here

Thus the wavelength of a given photon is proportional to a, and we can whimsically think of the corresponding electromagnetic waves as being "stretched out" at the same rate that space is being stretched out. Another way to interpret this redshifting is as a Doppler shift, since due to the expansion of the universe, the distant objects emitting the photons that we see are receding from us.

If we observe spectral lines in the spectrum of a distant object, they will all be shifted towards longer wavelengths by the same factor. Calling this factor $(1+z)$, the *redshift* of the object is defined to be the number z. For objects receding by a velocity $v \ll c$, we have to a good approximation that

$$z \approx \frac{v}{c}.$$

If we assume that the recession velocity v is mainly due to the expansion of the universe, we can immediately estimate the distance d to the object as

$$d \approx \frac{c}{H} z \approx 10^{28}\mathrm{cm} \times z.$$

(If z is of order unity or larger, this formula is replaced by a more complicated one.)

In cosmology, redshift is often used as a substitute for time. Indeed, the horizontal axis on most of the plots in the subsequent chapters is labeled z, not t. This is mainly because z corresponds much more directly to what astronomers actually measure. When we look out into space, we are looking into the past. If we observe a distant quasar, we know that we are observing events that occurred at some earlier time t, but we have no direct and accurate way of measuring t, or how long ago the quasar emitted the light that we see. However, we can measure its redshift simply by observing its spectrum and measuring where the spectral lines are. Thus "at redshift z" in cosmology jargon means "at the time when photons would have had to be emitted to have a redshift z by now." There is of course an explicit mathematical relationship between the two variables t and z, and this will be given later in this chapter. $z = \infty$ corresponds to $t = 0$ (the Big Bang), and $z = 0$ corresponds to today.

2.3.7 The Friedmann equation

So far, we have only discussed properties of the FRW metric. For it to correspond to something physical, it has to satisfy the Einstein field equations. If we take the stress-energy tensor $T_{\mu\nu}$ that corresponds to comoving matter of uniform density $\rho(t)$, substitute it into the Einstein field equations, and grind through a not particularly amusing calculation, it is found that the FRW metric is indeed a solution providing that the scale factor a satisfies the ordinary differential equation

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{k}{a^2}.$$

The solution to this equation, known as the *Friedmann Equation*[6], clearly depends both on the value of k and on the equation of state, the way in which the density ρ depends on a. If the density is dominated by nonrelativistic matter (particles whose kinetic energy is much less than their rest mass), just "matter" for short, it will obviously drop as the inverse of the volume as the universe expands, giving

$$\rho \propto a^{-3}.$$

This has been the case for quite a while, for redshifts $z \ll 10^4$. If the density is dominated by photons (or other highly relativistic particles), which become redshifted as discussed above,

$$\rho \propto a^{-4}$$

since the energy of each photon dwindles as $1/a$ as the universe expands. This was indeed the case very early on. A highly speculative third type of mass is *vacuum energy*, which if it exists has the property that its density stays constant, independent of the scale factor. The presence of such matter is equivalent to the presence of an infamous "cosmological constant" Λ in the Einstein field equations.

Let us define the *critical density*

$$\rho_{crit} \equiv \frac{3H^2}{8\pi G}$$

and the dimensionless variable

$$\Omega \equiv \frac{\rho}{\rho_{crit}}.$$

Then for the matter-dominated case (to which we limit ourselves in the remainder of this section), the Friedmann equation becomes

$$\frac{\dot{a}}{a} = H_0 \left[\Omega_0 \left(\frac{a}{a_0} \right)^{-3} + \lambda_0 - (\Omega_0 + \lambda_0 - 1) \left(\frac{a}{a_0} \right)^{-2} \right]^{1/2},$$

where if $k \neq 0$, we chose the scale factor today to be

$$a_0 \equiv \left(\frac{k}{\Omega_0 - 1} \right)^{1/2} H_0^{-1} \approx \frac{3000 h^{-1}\mathrm{Mpc}}{|\Omega_0 - 1|^{1/2}},$$

which can be interpreted as the current radius of curvature of the universe. If $\Omega_0 = 1$, then a_0 has no physical meaning and can be chosen arbitrarily — it will always cancel out when some physical quantity is computed. (Here and throughout cosmology, the subscript zero on a variable denotes its value today, at $z = 0$. For instance, H_0 is the value of the Hubble constant today.) λ_0 is the above-mentioned cosmological constant, believed by many to be zero. The constant k was eliminated by using the fact that

$$k = H_0^2(\lambda_0 + \Omega_0 - 1).$$

[6] This is an example where general relativity is in a sense more "classical" than special relativity. The Friedmann equation is identical to that describing the time-evolution of the radius a of a self-gravitating cloud of uniform, pressureless dust in classical physics. Note that c, the speed of light, appears nowhere in the equation, so with this classical interpretation, nothing prevents the dust particles from moving faster than the speed of light.

This relation shows that the universe is flat ($k = 0$) if $\lambda_0 + \Omega_0 = 1$, and closed or open depending on whether $\lambda_0 + \Omega_0$ exceeds unity or not. In the standard cold dark matter (CDM) scenario, one assumes that $\Omega_0 = 1$ and $\lambda_0 = 0$. For this simple case, the Friedmann equation has the solution

$$a(t) \propto t^{2/3},$$

and $\Omega = 1$ at all times. If the universe is open, it is clear that unless $\lambda_0 < 0$, we will always have $\dot{a} > 0$, and the expansion will continue forever. If the universe is closed and $\lambda_0 \leq 0$, the expansion eventually ceases and the universe contracts back together again, ending with a hot Big Bang in reverse, a big crunch. For the $\lambda_0 = 0$ case, the graph of the function $a(t)$ is a cycloid, and the Friedmann equation above describes only the increasing half of the curve. Closed universes expand forever if (Glanfield 1966)

$$\lambda_0 \geq 4\Omega_0 \cos^3\left[\frac{1}{3}\arccos\left(\Omega_0^{-1} - 1\right) + \frac{4\pi}{3}\right].$$

An recent review of properties of the Friedmann equation is given by Carroll *et al.* (1992).

2.3.8 Some handy formulas for getting by in curved space

This section contains a collection of useful formulas, many of which are used in the following chapters. It is intended as more of a reference section, and can be safely skipped by the general reader.

Friedmann solutions, conformal time

Since $a_0/a = 1 + z$, the Friedmann equation (neglecting radiation) can also be written

$$H = H_0\sqrt{\lambda_0 + (1+z)^2(1 - \lambda_0 + \Omega_0 z)}$$

or, alternatively,

$$-\frac{dz}{dt} = H_0(1+z)\sqrt{\lambda_0 + (1+z)^2(1 - \lambda_0 + \Omega_0 z)}. \tag{2.1}$$

In the rest of this section, we will set $\lambda_0 = 0$ unless the contrary is explicitly stated. This means that the Friedmann equation can be solved analytically, the $\Omega > 1$ solution being the ubiquitous cycloid.

It is often convenient to introduce a new time coordinate τ called *conformal time*, defined by

$$\tau(t) \equiv \int_0^t \frac{dt'}{a(t')}.$$

In terms of the conformal time, the solutions $a(t)$ to the Friedmann equation can be written in the parametric form

$$H_0 t = \begin{cases} \frac{\Omega_0}{2(1-\Omega_0)^{3/2}}(\sinh\tau - \tau) & \text{for } \Omega_0 < 1, \\ \frac{2}{3}\left(\frac{\tau}{\tau_0}\right)^3 & \text{for } \Omega_0 = 1, \\ \frac{\Omega_0}{2(\Omega_0-1)^{3/2}}(\tau - \sin\tau) & \text{for } \Omega_0 > 1, \end{cases}$$

$$\frac{a}{a_0} = \frac{1}{1+z} = \begin{cases} \frac{\Omega_0(\cosh\tau - 1)}{2(1-\Omega_0)} & \text{for } \Omega_0 < 1, \\ \left(\frac{\tau}{\tau_0}\right)^2 & \text{for } \Omega_0 = 1, \\ \frac{\Omega_0(1-\cos\tau)}{2(\Omega_0-1)} & \text{for } \Omega_0 > 1. \end{cases}$$

When $\Omega_0 \neq 1$, τ is often called the *development angle*. When $\Omega_0 = 1$, $\tau_0 = 2/(H_0 a_0)$, where the constant a_0 can be chosen arbitrarily. Again, this is because the radius of curvature, which fixed a_0 when $\Omega_0 \neq 1$, is infinite when the universe is flat. In all cases, $\tau = 0$ corresponds to the Big Bang. Inverting these last expressions, we obtain

$$\tau \equiv \begin{cases} 2\sinh^{-1}\sqrt{\frac{1-\Omega_0}{\Omega_0(1+z)}} & \text{for } \Omega_0 < 1, \\ \frac{\tau_0}{\sqrt{1+z}} & \text{for } \Omega_0 = 1, \\ 2\arcsin\sqrt{\frac{\Omega_0-1}{\Omega_0(1+z)}} & \text{for } \Omega_0 > 1. \end{cases}$$

Thus the conformal time today is

$$\tau_0 = \begin{cases} 2\sinh^{-1}\sqrt{\Omega_0^{-1} - 1} & \text{for } \Omega_0 < 1, \\ 2\arcsin\sqrt{1 - \Omega_0^{-1}} & \text{for } \Omega_0 > 1. \end{cases}$$

Although one cannot give an explicit expression for a in terms of t in general,

$$\frac{a}{a_0} \approx \Omega_0^{1/3}\left(\frac{3}{2}H_0 t\right)^{2/3}$$

for $t \ll H_0^{-1}$.

Causality, horizons, angles

A useful property of τ is that during a conformal time interval $d\tau$, a light ray always travels the same comoving distance $d\tau$. If $\Omega = 1$ so that space is flat, the situation becomes especially simple when viewed in $(\tau, \mathbf{x})$-space: all light rays (null geodesics) simply move in straight lines, and always make 45 degree angles with the τ axis. Thus after changing variables to comoving spatial coordinates and conformal time, we can use all our intuition about light propagation and causality in good old non-expanding Euclidean space. This is illustrated in Figure 2.1.

Parts of the universe receding from us faster than the speed of light are not in causal contact with us. No signals have ever been able to reach a comoving object in the universe from objects that are outside of its backward light cone. Such objects are said to be outside of its *horizon*. Thus defining the *horizon radius* as $a\tau$ and combining some of the previous expressions, we obtain

$$r_h \equiv a\tau = \begin{cases} \frac{H_0^{-1}}{(1+z)\sqrt{1-\Omega_0}}\cosh^{-1}\left[1 + \frac{2(1-\Omega_0)}{\Omega_0(1+z)}\right] & \text{if } \Omega_0 < 1, \\ \frac{2H_0^{-1}}{(1+z)^{3/2}} & \text{if } \Omega_0 = 1, \\ \frac{H_0^{-1}}{(1+z)\sqrt{\Omega_0-1}}\cos^{-1}\left[1 - \frac{2(\Omega_0-1)}{\Omega_0(1+z)}\right] & \text{if } \Omega_0 > 1. \end{cases}$$

If an object at redshift z has a proper diameter D (perpendicular to us), then one can show that it will subtend an angle on the sky given by

$$2\tan\frac{\theta}{2} = \begin{cases} \frac{\Omega_0^2(1+z)\sinh\left[\frac{1}{2}(1+z)\sqrt{1-\Omega_0}H_0 D\right]}{\sqrt{1-\Omega_0}\left[\Omega_0 z-(2-\Omega_0)(\sqrt{1+\Omega_0 z}-1)\right]} & \text{if } \Omega_0 < 1, \\ \frac{(1+z)^{3/2} H_0 D}{2(\sqrt{1+z}-1)} & \text{if } \Omega_0 = 1, \\ \frac{\Omega_0^2(1+z)\sin\left[\frac{1}{2}(1+z)\sqrt{\Omega_0-1}H_0 D\right]}{\sqrt{\Omega_0-1}\left[\Omega_0 z-(2-\Omega_0)(\sqrt{1+\Omega_0 z}-1)\right]} & \text{if } \Omega_0 > 1. \end{cases}$$

Thus if we look at an object at a redshift z, its horizon radius subtends an angle θ in the sky that is given by

$$2\tan\frac{\theta}{2} = \frac{\Omega_0^{3/2}\sqrt{1+z}}{\Omega_0 z - (2-\Omega_0)(\sqrt{1+\Omega_0 z}-1)},$$

for all values of Ω_0. For $\Omega_0 = 1$, this takes the particularly simple form

$$2\tan\frac{\theta}{2} = \frac{1}{\sqrt{1+z}-1},$$

whereas for large z, it reduces to

$$\theta \approx \sqrt{\frac{\Omega_0}{z}}.$$

This last expression is a rather poor approximation for for typical values of Ω_0, unless $z \gg 100$.

Age of the universe

The current age of the universe, t_0, is given by

$$t_0 = \int dt = \int_0^\infty (aH)^{-1}(z)\left|\frac{da}{dz}\right| dz = \int_0^\infty \frac{dz}{(1+z)H(z)}.$$

Performing this integral for $\lambda_0 = 0$ gives the age of the universe as

$$H_0 t_0 = \begin{cases} \frac{1}{1-\Omega_0} - \frac{\Omega_0}{2(1-\Omega_0)^{3/2}}\cosh^{-1}\left(\frac{2}{\Omega_0}-1\right) & \text{for } \Omega_0 < 1, \\ \frac{2}{3} & \text{for } \Omega_0 = 1, \\ \frac{\Omega_0}{2(\Omega_0-1)^{3/2}}\cos^{-1}\left(\frac{2}{\Omega_0}-1\right) - \frac{1}{\Omega_0-1} & \text{for } \Omega_0 > 1. \end{cases}$$

For the flat case where $k = 0$ and $\Omega_0 = 1 - \lambda_0$, the result is

$$H_0 t_0 = \frac{1}{\lambda_0^{1/2}}\ln\left[\frac{1+\lambda_0^{1/2}}{(1-\lambda_0)^{1/2}}\right].$$

2.4 Perturbing the Universe

In this section, we will give the rudiments of the theory of structure formation. First we discuss how small ripples in the FRW metric grow larger as the universe expands. We then briefly mention the theory of Gaussian random fields and the so called Press-Schechter model for galaxy formation.

2.4.1 Linear perturbation theory

Let us assume that demand is an arbitrary function of price; $D = aP + b$.
Anonymous economics professor

In this section, everything will indeed be linear. As mentioned in the previous section, it is generally futile to look for analytic solutions to the Einstein field equations of GR. Fortunately, the actual metric is fairly uniform on large scales, indicating that the relative deviations from the FRW metric are rather small. For this reason, cosmologists attack the problem of structure formation with linear perturbations. One perturbs the FRW metric, throws away all terms of second order and higher, and after Fourier transforming ends up with a simple equation for the time-evolution of the various perturbation modes.

As mentioned in a footnote a few pages back, the Friedmann equation can be derived using classical physics alone. This is no coincidence. A result known as Birkhoff's theorem states that if we take an FRW metric and remove all matter from a spherical volume, then space will be flat inside of this sphere, *i.e.* all gravitational effects from the matter outside cancel out. If we put the uniformly expanding matter back into the sphere, then its gravitational behavior will be correctly described by classical physics. As long as the gravitational field is weak (which it is in the present problem, as opposed to say near a black hole) and the sphere is much smaller than the horizon scale 10^{10} light years, it is easy to show that classical physics will apply even if the matter is not uniform. With this motivation, let us write down the classical equations governing a gas of density ρ, velocity $\mathbf{v}$ and pressure p and the corresponding gravitational potential ϕ. These are the continuity equation, the Euler equation of motion and the Poisson equation of classical gravity, respectively:

$$\begin{cases} \dot{\rho} + \nabla \cdot (\rho \mathbf{v}) = 0 \\ \dot{\mathbf{v}} + (\mathbf{v} \cdot \nabla)\mathbf{v} = -\nabla(\phi + \frac{p}{\rho}) \\ \nabla^2 \phi = 4\pi G \rho \end{cases} \tag{2.2}$$

A simple solution to these equations is

$$\begin{cases} \rho_0(t, \mathbf{r}) = \frac{\rho_0}{a^3}, \\ \mathbf{v}_0(t, \mathbf{r}) = \frac{\dot{a}}{a}\mathbf{r}, \\ \phi_0(t, \mathbf{r}) = \frac{2\pi G \rho_0}{3} r^2, \end{cases}$$

where a is a function of time that satisfies the Friedmann equation, and since the density is independent of $\mathbf{r}$, this classical solution clearly corresponds to the unperturbed FRW solution in GR. From here on, let us write these fields as functions of the *comoving* position

$\mathbf{x}$ rather than the physical position $\mathbf{r} = a(t)\mathbf{x}$. Let us now expand the fields as

$$\begin{cases} \rho(t,\mathbf{x}) = \rho_0(t)[1+\delta(t,\mathbf{x})], \\ \mathbf{v}(t,\mathbf{x}) = \mathbf{v}_0(t,\mathbf{x}) + \mathbf{v}_1(t,\mathbf{x}), \\ \phi(t,\mathbf{x}) = \phi_0(\mathbf{x}) + \phi_1(t,\mathbf{x}), \end{cases}$$

and assume that $|\delta|$, $|\mathbf{v}_1|$, $\phi_1 \ll 1$. Substituting this into equation (2.2), dropping all non-linear terms, Fourier transforming everything with respect to $\mathbf{x}$, and doing some algebra, one obtains the second order ordinary differential equation

$$\ddot{\hat{\delta}}(\mathbf{k}) + 2\frac{\dot{a}}{a}\dot{\hat{\delta}}(\mathbf{k}) + \left(\frac{v_s^2|\mathbf{k}|^2}{a^2} - 4\pi G\rho_0\right)\hat{\delta}(\mathbf{k}) = 0. \tag{2.3}$$

Here the equation of state for the gas enters only through the *sound speed*,

$$v_s \equiv \left(\frac{\partial p}{\partial \rho}\right)_{adiab}^{1/2} \approx \sqrt{\frac{5kT}{3m_p}},$$

where the last expression applies if the gas consists mainly of atomic hydrogen (m_p denotes the proton mass). If we substitute an expanding universe solution for a, equation (2.3) tells us a number of interesting things.

The coefficient of $\hat{\delta}$ has two contributions with opposite signs: the destabilizing effect of gravity competes with the stabilizing effect of gas pressure. On small scales (corresponding to large k), pressure dominates and perturbations do not grow, merely oscillate. On large scales, on the other hand, self-gravity dominates, and fluctuations grow monotonically until δ becomes of order unity and the entire linear approximation breaks down. The border between these two regimes is when the perturbation has a physical wavelength

$$\lambda \equiv \frac{2\pi a}{|\mathbf{k}|},$$

called the *Jeans wavelength*

$$\lambda_J \equiv v_s\sqrt{\frac{\pi}{G\rho_0}},$$

for which this coefficient vanishes.

The coefficient of $\dot{\hat{\delta}}$, twice the Hubble constant H, acts like a kind of friction: in the absence of this term, we could get exponentially growing modes, but as will be seen, the growth is never that rapid in the presence of this term, *i.e.* in an expanding universe.

If the density field is smoothed on a scale $\lambda \gg \lambda_J$, then we can neglect the k-dependence and trivially Fourier transform back to real space. Thus all contributing modes grow at the same rate, and a perturbation retains its (comoving) shape as the universe expands, simply increasing in amplitude. For the simple case where $\Omega = 1$, $\lambda_0 = 0$ and $a \propto t^{2/3}$, we obtain

$$\ddot{\delta} + \frac{4}{3t}\dot{\delta} - \frac{2}{3t^2}\delta = 0,$$

which has the most general solution

$$\delta(t,\mathbf{x}) = A(\mathbf{x})\delta(t_0,\mathbf{x})\left(\frac{t}{t_0}\right)^{2/3} + B(\mathbf{x})\delta(t_0,\mathbf{x})\left(\frac{t}{t_0}\right)^{-1}$$

for some time-independent functions A and B. Ignoring the physically uninteresting decaying mode, and writing δ as a function of redshift rather than time, we thus have the extremely simple result

$$\delta(z, \mathbf{x}) \propto \frac{\delta_0(\mathbf{x})}{1+z},$$

where δ_0 is the linearly extrapolated field of density perturbations today.

2.4.2 Random fields

The mathematical theory of random fields (sometimes known as three-dimensional stochastic processes) is a very useful tool when analyzing cosmological structure formation. A *random field* is simply an infinite-dimensional random variable, such that each realization of it is a field on some space. In cosmology applications, these fields tend to be δ, $\mathbf{v}$ or ϕ, and the space is physical space at some fixed time t. As every quantum field theorist knows, it is a nightmare to try to define a nice measure on an infinite-dimensional space, so random fields are defined by specifying the joint probability distribution of their values at any n points, $n = 1, 2, 3, ...$, thus circumventing the need to define a probability distribution on the space of all fields. Hence, to define a random field δ, one must specify the 1-point distribution (the 1-dimensional probability distribution of $\delta(\mathbf{x}_1)$ for all $\mathbf{x}_1$), the 2-point-distribution (the 2-dimensional probability distribution of the vector $[\delta(\mathbf{x}_1), \delta(\mathbf{x}_2)]$ for all $\mathbf{x}_1$ and $\mathbf{x}_2$), *etc.* In cosmology, the random fields are always assumed to be translationally and rotationally invariant, *i.e.* homogeneous and isotropic. Hence the 1-point distribution is independent of $\mathbf{x}_1$, and the 2-point distribution will depend only on the scalar quantity $x \equiv |\mathbf{x}_2 - \mathbf{x}_1|$.

Ergodicity

A random field δ is said to be *ergodic* if one can use ensemble averaging and spatial averaging interchangeably. The ensemble average of a random field δ at a point, denoted $\langle\delta(\mathbf{x})\rangle$, is simply the expectation value of the random variable $\delta(\mathbf{x})$. Thus for an ergodic field,

$$\langle\delta(\mathbf{x}_1)\rangle = \lim_{R\to\infty}\left(\frac{4}{3}\pi R^3\right)^{-1}\int_{|\mathbf{x}|<R}\delta_*(\mathbf{x})d^3x$$

holds for all points $\mathbf{x}_1$ and for all realizations $\delta_*(\mathbf{x})$ (except for a set of probability measure zero). Ensemble averages are completely inaccessible to us, since we have only one universe to look at, namely the particular realization that we happen to live in. So as cosmologists, we are quite happy if we have ergodicity, since this means that we can measure these elusive ensemble averages by simply averaging over large volumes instead.

Gaussianity

A random field is said to be *Gaussian* if all the above-mentioned probability distributions are Gaussian. This is a very popular assumption in cosmology, partly because, as we will see, it greatly simplifies matters. A first nice feature of this assumption is that all homogeneous and isotropic Gaussian random fields are ergodic[7]. Taking the spatial average of the definition of δ, for instance, ergodicity implies that

$$\langle\delta(\mathbf{x})\rangle = 0.$$

Let us define the r.m.s. fluctuations

$$\sigma \equiv \langle\delta(\mathbf{x})^2\rangle^{1/2},$$

and the *correlation function* as

$$\xi(x) \equiv \frac{\langle\delta(\mathbf{x}_2)\delta(\mathbf{x}_1)\rangle}{\sigma^2}.$$

(Note that because of the homogeneity and isotropy, σ is independent of $\mathbf{x}$ and the correlation function depends only on $x \equiv |\mathbf{x}_2 - \mathbf{x}_1|$.) Since the n-point distribution is Gaussian, it is defined by its mean vector $\langle\delta(\mathbf{x}_n)\rangle$ (which is identically zero) and its covariance matrix

$$C_{mn} \equiv \langle\delta(\mathbf{x}_m)\delta(\mathbf{x}_n)\rangle = \sigma^2\xi(|\mathbf{x}_m - \mathbf{x}_n|).$$

Thus the Gaussian random field δ has the extremely useful property that it is is entirely specified by σ and its correlation function.

The power spectrum

If we Fourier expand δ as

$$\delta(\mathbf{x}) = \frac{1}{(2\pi)^3}\int e^{i\mathbf{k}\cdot\mathbf{x}}\widehat{\delta}(\mathbf{k})d^3k,$$

we see that if its Fourier transform

$$\widehat{\delta}(\mathbf{k}) = \int e^{-i\mathbf{k}\cdot\mathbf{x}}\delta(\mathbf{x})d^3x,$$

is a Gaussian random variable for any k, then δ will automatically be a Gaussian random field (since a sum of Gaussians is always Gaussian). Cosmologists like to postulate that the complex numbers $\widehat{\delta}(\mathbf{k})$ have *random phases*, which implies that they are all uncorrelated. One postulates that

$$\langle\widehat{\delta}(\mathbf{k})^*\widehat{\delta}(\mathbf{k}')\rangle = (2\pi)^3\delta(\mathbf{k}-\mathbf{k}')P(\mathbf{k}),$$

where the function $P(\mathbf{k})$ is called the *power spectrum* and δ is the Dirac delta function (which will hopefully not be confused with the random field δ). This implies that even if $\widehat{\delta}(\mathbf{k})$ does not have a Gaussian distribution, the random field δ, being an infinite sum of independent random variables, will still be Gaussian by the Central Limit Theorem for many well-behaved power spectra. It is straightforward to show that the power spectrum is simply σ^2 times the three-dimensional Fourier transform of the correlation function, *i.e.*

$$P(k) = 4\pi\sigma^2\int\left(\frac{\sin kr}{kr}\right)\xi(r)r^2dr.$$

Note that P depends on $\mathbf{k}$ only through its magnitude $k = |\mathbf{k}|$, because of the isotropy assumption.

[7] Note that this is true only for random fields that live on infinite spaces such as R^n, and does not hold for fields on compact manifolds such as the sphere S^2. Thus the field of microwave background anisotropies (defined further on) is not ergodic, so that even if we could reduce our experimental errors to zero, we could still never measure any ensemble-averages with perfect accuracy. This phenomenon is known as "cosmic variance". It stems from the fact that in the spatial average above, one cannot average over an infinite volume (*i.e.* let $R \to \infty$), since the volume of the compact manifold is finite.

Window functions

If we smooth δ by defining the weighted average

$$\delta_w(\mathbf{x}) \equiv \int \delta(\mathbf{r}')w(\mathbf{r}'-\mathbf{r})d^3r',$$

w being some weight function, then

$$\sigma_w^2 \equiv \langle|\delta_w(\mathbf{x})|^2\rangle = \frac{1}{(2\pi)^3}\int|\widehat{w}(k)|^2P(k)d^3k.$$

This is quite a typical pattern: the variance of some physical quantity is given by an integral of the power spectrum against some function, referred to as a *window function.* In this particular example, where for an appropriately chosen w the physical quantity can be interpreted as the mass within a volume of some given shape, the window function is $|\widehat{w}|^2$. As we will see in Chapter 7, basically all cosmological experiments that measure a single number can be described by some window function, regardless of whether they involve mass distribution, bulk flows of galaxies or microwave background fluctuations.

2.4.3 The Press-Schechter recipe

In this section, we will describe a rather simple-minded model of structure formation that agrees remarkably well with observational data.

The top hat solution

Early on, while $z \gg \Omega_0^{-1}$, space is approximately flat and the Friedmann equation has the approximate solution

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

regardless of the values of Ω_0 and λ_0. If an $\Omega = 1$ universe has a completely uniform density ρ except for a "top hat" overdensity, a spherical region where the density at a given time is some constant $\rho' > \rho$, then this top hat region will gradually begin to expand slower than the rest of the universe, stop expanding and recollapse to a point. By Birkhoff's theorem, the radius of this region will evolve according to the Friedmann equation, but with $\Omega > 1$. It is easy to show that the overdensity

$$\delta \equiv \frac{\rho'}{\rho} - 1$$

evolves as

$$(1+\delta) = \frac{9}{2}\frac{(\alpha-\sin\alpha)^2}{(1-\cos\alpha)^3} = 1 + \frac{3}{20}\alpha^2 + O(\alpha^3), \tag{2.4}$$

where the parameter α, the "development angle", is related to the redshift through

$$\frac{1+z_{vir}}{1+z} = \left(\frac{\alpha-\sin\alpha}{2\pi}\right)^{2/3} = \frac{\alpha^2}{(12\pi)^{2/3}} + O(\alpha^{8/3}). \tag{2.5}$$

Here z_{vir} is the redshift at which the top hat would collapse to a point. In reality, an overdense region would of course not collapse to a point (and form a black hole). Since it would not be perfectly spherically symmetric, dark matter particles would mostly miss each other as they whizzed past the central region and out again on the other side, eventually settling down in some (quasi-) equilibrium configuration known as a virial state. For baryons, gas-dynamical processes become important, and pressure eventually slows the collapse. Strictly speaking, virial states are not stable over extremely long periods of time, and their density is certainly not uniform. For a virialized lump, often referred to as a "halo", a typical density profile peaks around some constant value in its core and falls off like $1/r^2$ over some range of radii. Nonetheless, people often say that they have a "typical" density

$$\rho_{vir} \approx 18\pi^2\rho_0(1+z_{vir})^3,$$

which is a useful rule of thumb. Thus in the top-hat collapse model, density in the perturbed region is assumed to evolve as in Figure 2.2. The density starts out decreasing almost as fast as the background density ρ, with

$$\delta \propto (1+z)^{-1}$$

early on, just as in linear theory, but gradually stops decreasing and increases radically as z approaches z_{vir}. It never increases past the virial value $\approx 18\pi^2\rho_0(1+z_{vir})^3$, however, but stays at that density for all $z < z_{vir}$. Unfortunately, α cannot be eliminated from the equations that relate δ and z, *i.e.* equations (2.4) and (2.5), by using elementary functions. For this reason, I made the following fit to the function $\delta(z)$, which is accurate to about 5% until z is within 10% of z_{vir}, at which the density is assumed to start approaching the virial value anyway:

$$\rho(z) \approx \rho_0(1+z)^3 \exp\left[-\frac{1.9A}{1-0.75A^2}\right],$$

where

$$A(z) \equiv \frac{1+z_{vir}}{1+z}.$$

The Press-Schechter formula

In the above top-hat solution, we have approximately

$$\delta = \frac{3}{20}(12\pi)^{2/3}A \approx 1.686A$$

early on, while $A \ll 1$. If we naively extrapolate this far beyond its region of validity, we thus obtain

$$\delta = \delta_c \equiv \frac{3}{20}(12\pi)^{2/3} \approx 1.69$$

when $A = 1$, *i.e.* when $z = z_{vir}$. Press and Schechter proposed the following simplistic prescription for modeling structure formation: define a smoothed density field δ_M by averaging δ over spherical regions of such a radius that they contain a mass M on the average. Define $\sigma(M)$ as the r.m.s. mass fluctuation in these spheres, *i.e.*

$$\sigma(M) \equiv \langle\delta_M^2\rangle^{1/2}.$$

Now make the interpretation that a point $\mathbf{x}$ is part of a virialized halo of mass exceeding M if

$$\delta_M(\mathbf{x}) \geq \delta_c \approx 1.69.$$

In linear theory, for $\Omega_0 = 1$ and $\lambda_0 = 0$, we clearly have

$$\sigma(M,z) = \frac{\sigma(M,0)}{1+z}.$$

Since δ is a Gaussian random field, so is δ_M, which means that $\delta_M(\mathbf{x})$ is a Gaussian random variable with zero mean and a standard deviation $\sigma(M,z)$. Thus with the Press-Schechter interpretation, the fraction of all mass that is in virialized lumps of mass exceeding M at a redshift z is

$$f_g(M,z) \equiv P[\delta_M(\mathbf{x}) > \delta_c] = \frac{1}{2}\mathrm{erfc}\left[\frac{(1+z)\delta_c}{\sqrt{2}\sigma(M,0)}\right],$$

where the complementary error function is defined by

$$\mathrm{erfc}[x] \equiv \frac{2}{\sqrt{\pi}}\int_x^\infty e^{-x^2}\,dx.$$

Unfortunately, the so defined $f_g(M,z)$ approaches $\frac{1}{2}$ as $t \to \infty$, *i.e.* half of all mass remains unaccounted for. To remedy this, Press and Schechter introduced a "fudge factor" of 2 and predicted that

$$f_g(M,z) = \mathrm{erfc}\left[\frac{(1+z)\delta_c}{\sqrt{2}\sigma(M,0)}\right].$$

And lo and behold: this is quite a good fit both to the observed mass distribution of galaxies today and to the $f_g(M,z)$ that is observed in numerical n-body simulations. Although a score of papers have been written on the topic since this formula was proposed some twenty years ago, often attempting to justify the factor of two, not much has happened to the basic formula except that other values have been proposed for the constant δ_c, ranging from 1.33 to 2.00.

2.4.4 CDM and BDM

After all this talk about the power spectrum $P(k)$, what does theory predict it to be? There is now a plethora of rival theories on the market predicting all sorts of power spectra that are tailor-made to fit various observational facts. In this extremely cursory discussion, we will give only the highlights of two models. The most popular model over the last decade has probably been the cold dark matter (CDM) model with adiabatic initial fluctuations and an $n=1$ Harrison Zel'dovich initial power spectrum. We will refer to this mouthful as "standard CDM" for short. This model is simple and elegant, as it contains almost no free parameters, and roughly speaking fits all the observed power spectrum data to within about a factor of two when k ranges over many orders of magnitude. A recent rising star (albeit always surrounded by rumors of its imminent death) is the baryonic dark matter (BDM) model with isocurvature initial fluctuations and a power-law initial power spectrum. We will refer to this as just BDM for short.

All theories of structure formation (theories that predict $P(k)$) split into two parts: one part that gives some "initial" or "primordial" power spectrum $P_i(k)$ when the universe was very young, and a second part that describes how this power spectrum is altered by various physical processes to produce the $P(k)$ we observe today. Neglecting speculative theoretical constructs such as cosmic strings and other so called topological defects, the latter processes usually do not mix different Fourier modes as long as $\delta \ll 1$. Thus the second half of a theory can be specified by a single function, the *transfer function* $T(k)$, defined so that

$$P(k) = T(k)^2 P_i(k).$$

The initial power spectrum

For lack of any better assumptions, cosmologists usually assume that the primordial power spectrum was a simple power law

$$P(k) \propto k^n$$

for some constant n termed the *spectral index*, at least over the range of k-values that are accessible to our observations. In CDM, the primordial fluctuations are usually assumed to arise from quantum fluctuations during a speculative inflationary epoch when the universe was less than 10^{-32} seconds old. Thirteen years after the theory of inflation was invented by Alan Guth, there are now scores of different versions of it. Most of them predict a spectral index between 0.7 and 1. These initial fluctuations are *adiabatic*, which means that the density varies slightly from point to point but that the ratio of photons to baryons and dark matter remains constant initially. For BDM, the primordial spectrum is an Achilles' heel, as the theory contains no physical mechanism for generating it. One simply postulates that it is a power-law. In addition, one postulates that the initial fluctuations are of *isocurvature* type, which means that the fluctuations in the photon density cancel the baryon fluctuations, so that the total density remains constant. One may question the value or calculating the BDM transfer function, since at the end of the day it is multiplied by a function that was put in arbitrarily by hand, but peoples' opinions differ on what is elegant. An aesthetic advantage of BDM is that it requires no new particles besides those we know exist, whereas CDM postulates that more than 90% of the matter is of a type that we have never detected in our laboratories.

The CDM transfer function

The standard way to compute transfer functions is to write down the Boltzman equation for the time-evolution of the phase-space distributions of all particle species in the model (baryons, photons, neutrinos, collisionless dark matter, *etc.*), linearize it by throwing away all quadratic and higher order terms, plug it into a fast computer and go for lunch. To model evolution of the power spectrum at late times, when the approximation $\delta \ll 1$ breaks down, so called N-body simulations are employed: the time-evolution of some 10^6 to 10^8 point-particles in a box is studied by numerically integrating the Newtonian equations of motion, assuming periodic boundary conditions. The power spectra resulting from a few assorted models are plotted in Figure 2.3.

Fortunately, the main features of the CDM spectrum can be understood without numerical calculations, with a bit of hand-waving, as we will now see. If the transfer function is to be anything else than a simple power law in k, some process must imprint a physical length scale on it. Collisionless dark matter particles have no particular length scale at all associated with them that is relevant in this context. However, at redshifts $z \gtrsim 10^4$, the main contribution to the density and the self-gravity of the universe was CBR photons, not dark matter, so before this, the crucial issue was what the photons decided to do. Were it not for this fact, a power-law initial power spectrum would necessarily lead to a power-law spectrum today. In fact, the transfer function would be simply $T(k) = 1$, giving

$$P(k) \propto k^n.$$

As the universe expands, the horizon size keeps increasing, also in comoving coordinates. This means that the wavelength of any sinusoidal perturbation (a distance which remains constant in comoving coordinates) will sooner or later become smaller than the horizon scale. When this happens, the perturbation is said to "enter the horizon". To avoid various painful GR gauge ambiguities associated with super-horizon sized modes, it is convenient to specify the amplitude of perturbations with a given wavenumber k when they enter the horizon. Since microphysical processes such as pressure support turn out to be irrelevant for super-horizon sized modes, the power spectrum at horizon entry must also have been some power law.

Before recombination, the photons and baryons were tightly coupled and acted as a single fluid where the sound speed was of the order of the speed of light. This means that the above-mentioned Jeans length was enormous, indeed several times greater than the horizon size, so perturbations that had entered the horizon did not grow much at all early on, indeed logarithmically at best. As the redshifting of the photons caused the density of the universe to become matter-dominated around $z \approx 10^4$, the perturbations in the dark matter began to grow. And at $z \approx 10^3$, recombination caused the Jeans length to drop drastically, down to a comoving scale much lower than the sizes of present-day galaxies, and perturbations in the baryons began to grow as well. Thus dark matter fluctuations that entered the horizon long after matter-radiation equality, fluctuations whose comoving wavelength exceeds

$$\lambda_{eq} \approx 13(\Omega_0 h^2)^{-1}\mathrm{Mpc},$$

were never affected by any of this microphysics that has a length scale associated with it, which means that $T(k) \approx 1$ for $k \ll \lambda_{eq}^{-1}$. Fluctuations with $k \gg \lambda_{eq}^{-1}$, on the other hand, entered the horizon while dark matter was still only a minuscule contributor to the overall density, and did not start growing properly until after matter-radiation equality. Thus we know that for $k \gg \lambda_{eq}^{-1}$, $T(k)^2$ must be some different power law k^m, with $m < 0$. We will now see that $m = -4$. Take $\Omega = 1$ and consider a perturbation of comoving wavelength λ. Then its physical scale is

$$d_\lambda = \frac{\lambda}{1+z}$$

whereas the horizon scale is

$$d_H \propto t \propto (1+z)^{-3/2}.$$

Equating these two scales gives the redshift for horizon entry

$$(1 + z_{ent}) \propto \lambda^{-2} \propto k^2.$$

Since this redshift, δ has grown by a factor of $(1+z_{ent})$, so $P(k)$ has grown by a factor of $(1+z_{ent})^2 \propto k^4$. The modes with $k \gg \lambda_{eq}^{-1}$, on the other hand, have all grown by the same factor since they entered the horizon, so in this part of the power spectrum, the slope is down by the sought for factor of k^4. In conclusion, the net result of all this is that

$$P(k) \propto \begin{cases} k^n & \text{for } k \ll \lambda_{eq}^{-1}, \\ k^{n-4} & \text{for } k \gg \lambda_{eq}^{-1}, \end{cases}$$

where $\lambda_{eq} \approx 50\,\mathrm{Mpc}$ for $\Omega = 1$ and $h = 0.5$.

The transfer functions for baryons and photons

All this concerned the power spectrum of the density of dark matter. The baryons and the photons are affected by two additional effects, known as Landau damping and Silk damping, which for the case of adiabatic fluctuations effectively wipes out their fluctuations on galactic scales. Since we know that galaxies do indeed exist, a successful theory of structure formation must find a way around this disaster. With CDM, the trick is that the cold dark matter is unaffected by this damping, and that gravitational instability eventually makes the baryon fluctuations "catch up" with the fluctuations in the dark matter. With BDM, the trick is to utilize isocurvature fluctuations instead, which do not suffer Silk damping. We will return to the transfer function for photons in the CBR section.

The complaint department

As to which model of structure formation is correct, the jury is still out. We will conclude this section by criticizing all the models, as democratically as possible. According to many workers, standard CDM is off by about a factor of two. This may not be much to fuss about in a field where one is often satisfied with getting the right answer to within an order of magnitude, but as data gets better, concern grows. This factor of two can be dealt with in at least six ways:

- One can argue that what we see is a biased version of what is actually there, and introduce different fudge factors called *bias factors* for various types of measurements.

- The spectral index n can be lowered to around 0.7 by tweaking certain parameters associated with inflation, which seems to underproduce high redshift quasars. So do several of the other models listed below.

- Repeating what Einstein described as "the biggest blunder of my life" (Gamow 1970, p. 44), a non-zero cosmological constant Λ can be introduced.

- Yet another free parameter can be introduced by adding just the right amount of massive neutrinos (about 30%) to the cosmological mixture.

- λ_{eq} can be increased by postulating that the τ neutrino decays in an appropriate manner and that there are new and unseen particles, for instance additional massless Nambu-Goldstone bosons.

- λ_{eq} can be increased by lowering the Hubble constant to about $h = 0.25$, causing raised eyebrows among those astronomers who claim that $h = 0.8$.

In addition, all of these inflation-based theories are challenged by the fact that some CBR experiments have been argued to indicate $n \approx 1.5$ on large scales, whereas essentially all inflationary scenarios predict $n \leq 1$.

Meanwhile, BDM is fighting a difficult battle, ensnarled by a number of tight constraints. The Compton y-parameter prevents reionization from being too complete, too hot and too early (see Chapter 5), degree-scale CBR-experiments prevent reionization from being too late, and the observed power-spectrum constrains the thermal history as well. As the author is writing this, one of his colleagues is running a Boltzmann code for BDM models which have nine free parameters, so it appears that in dodging the many constraints, the BDM model has lost much of its initial elegant simplicity, and with that perhaps most of its appeal.

2.5 The Microwave Background

In 1964, Arno Penzias and Robert Wilson at Bell Laboratories discovered microwave radiation of an unknown source that they at one point thought might have been bird droppings in their antenna horn. Eventually, their discovery was given a more grandiose interpretation, which resulted in them both winning a free trip to Stockholm. This radiation is the topic of the present section. In the first few subsections, where its origin is described, we will also make a digression about the thermal history of the universe.

2.5.1 The thermal history of the universe

When electrons are accelerated, they emit photons. Thus if one heats a hydrogen plasma sufficiently, the fast-moving electrons will frequently scatter off of nuclei and other electrons, resulting in a significant number of photons being created and radiated away. For cosmological applications, the two most important of these scattering processes are thermal bremsstrahlung

$$p^+ + e^- \mapsto p^+ + e^- + \gamma$$

and double Compton scattering

$$e^- + \gamma \mapsto e^- + \gamma + \gamma.$$

These processes can also go in the reverse direction, so if such a hot plasma filled the entire universe, some of these photons would eventually get into collisions where they would become absorbed. One can show that if the plasma is held at a constant, uniform density long enough, then a thermal equilibrium situation will arise where the electrons and protons have Maxwell-Boltzman velocity distributions corresponding to some temperature T_e and the photons have a Planck blackbody distribution

$$n_\gamma(\mathbf{x}, \mathbf{k}) = \frac{1}{e^{h\nu/kT_\gamma} - 1},$$

where $T_\gamma = T_e$. Here the photon frequency $\nu = |\mathbf{k}|c/2\pi$, and the wave vector $\mathbf{k}$ is not to be confused with Boltzmann's constant k. Note that since the number density n_γ is a distribution over the six-dimensional photon phase space $(\mathbf{x}, \mathbf{k})$, it is dimensionless, as the units of m^{-3} arising from $\mathbf{x}$ are canceled by the m^3 arising from $\mathbf{k}$.

In the beginning, there was thermal equilibrium

This was not quite the situation in the early universe, since the expansion kept lowering the density (and hence the reaction rates). However, it is easy to show that there was nonetheless sufficient time for thermal equilibrium to be attained early on. Thus at high redshifts (say $z \gg 10^3$), the entire state of the hydrogen plasma and the photons could be described by a single number, the temperature $T = T_e = T_\gamma$. Indeed, if we go back to high enough redshifts, thermal equilibrium was attained not only for electron and proton velocities and for electromagnetic radiation, but indeed for all particles that we know of, as a plethora of reactions kept rapidly creating and destroying all sorts of particles left and right.

For instance, when the universe was much less than a second old, electrons and positrons where being created and destroyed so rapidly that their densities where accurately given by the standard Boltzman formula. And even earlier on, the same applied to heavier particles like protons and W bosons.

Freezeout

As the universe expanded, the reaction rates that affected the abundance and energy distribution of any given particle species kept dropping until that species "decoupled" from the merry thermal mixture, or "froze out" as cosmologists say[8]. One by one, species after species froze out. If the rest mass of a particle was much greater than the thermal energy kT when its species froze out, then tough luck: the freezeout abundance would have been exponentially suppressed by a huge Boltzman factor, and almost no such particles would still be around today to bear witness. This was the grim fate of the W^+, W^- and Z bosons, which constitute only a minuscule fraction of the density of the universe today. This is also the assumed fate of many highly speculative particles[9] whose existence today would be in conflict with observations. In the other extreme, where the particle is so weakly interacting that it freezes out while kT is still larger than its rest mass, there is still an ample abundance of such particles today. Such weakly interacting massive particles, known as WIMPS, are interesting candidates for the dark matter in the universe. Another classic example of freezeout survivors are the cosmic background neutrinos, which are still with us today. A most interesting intermediate case is the neutron. Being about 0.2% heavier than the proton, the neutron to proton ratio in equilibrium is

$$\frac{n_n}{n_p} \approx e^{-10^{10}\mathrm{K}/T}.$$

Thus the neutron and proton abundances would be almost identical while the temperature of the universe exceeded their energy difference, but eventually, as $T \to 0$, the ratio would approach zero exponentially. Thus we might expect to find either virtually no neutrons today or a 50% abundance. In the latter case, it turns out that virtually all the hydrogen would have formed Helium during nucleosynthesis, and we would find things such as water quite difficult to come by. If we were around... Interestingly, as mentioned in the nucleosynthesis section above, nature chose to be right in the middle: neutron freezeout occurred just around this critical temperature, and the primordial abundance is about 12%.

After this digression, we will limit ourselves entirely to electrons and photons.

[8]a more quantitative definition is to say that a given reaction is frozen out when the rate Γ at which a particle is subject to such reactions is so low, that less than one such event is expected to occur while the universe doubles its age. This happens roughly when $\Gamma = H$, the Hubble constant.

[9]Particle physicists desperate for funding have postulated the existence of many dozens of particles never seen by man nor beast...

Recombination and the CBR freezeout

The freezeout picture applies to hydrogen recombination as well. In thermal equilibrium, the fraction of hydrogen that is ionized is

$$x \approx \left[1 + 2.8 \times 10^{-6} T_5^{-7/6} e^{1.58/T_5}\right]^{-1},$$

where $T_5 \equiv T/10^5\mathrm{K}$, which means that if one slowly lowers the temperature of a hot hydrogen plasma, it will recombine when $T \approx 15,000\mathrm{K}$. However, the recombination reaction

$$p^+ + e^- \mapsto H$$

has already begun freezing out by then. Quantum effects such as stimulated emission complicate the process further, and the end result is that recombination does not occur until much later, at a temperature of a few thousand K, at $z \approx 10^3$. At this point, something else of great importance happens as well: the photons freeze out. Until then, they were constantly Thomson scattering off of free electrons, but from this point on, a generic photon merely moves along in a straight line (or, more precisely, a geodesic) until today. Another way of phrasing this is that the universe became transparent at $z \approx 10^3$, so that in a sense we can see that far back. These frozen out photons gradually become redshifted as was described in the GR section, and it is easy to show that this redshifting preserves the blackbody nature of their spectrum, simply lowering the blackbody temperature as

$$T_\gamma = (1 + z) T_{\gamma 0}$$

as the the universe expands. It is these frozen out photons that are known as the cosmic microwave background radiation, the CBR. The best measurement to date to puts their current temperature at

$$T_{\gamma 0} \approx 2.726\mathrm{K} \pm 0.005\mathrm{K},$$

and was made by the Far Infra-Red Absolute Spectrometer (FIRAS) detector on board the celebrated COBE satellite (Mather *et al.* 1994).

2.5.2 CBR fluctuations

If one corrects for our motion relative to the comoving CBR rest frame[10], the observed CBR is almost perfectly thermal and isotropic. In other words, its spectrum is almost exactly a Planck spectrum, and the temperature $T_{\gamma 0}$ is almost exactly the same in all directions in the sky. If it were not for the caveat "almost", the CBR would not be so interesting, as it would contain no more information than the single number $T_{\gamma 0}$. (In addition, a perfectly isotropic CBR would have implied that all our theories about structure formation would be dead wrong, as we shall see below.) It is the small deviations from thermality and isotropy that have the potential to give us information about the early universe.

[10] Note that whereas there are no preferred observers in classical physics or in generic GR metrics, which renders concepts such as "at rest" meaningless, this is in fact not the case in the FRW metric. Here comoving observers have a special status, and the CBR plays the role of the infamous "aether" which Michelson and Morley (1887) discredited with their famous speed-of-light experiment.

These deviations can be conveniently split into two categories: spectral fluctuations and spatial fluctuations. To study the former, we average the spectrum from all directions in the sky and investigate the extent to which it deviates from a Planck spectrum. To study the latter, we summarize the spectrum in each direction $\hat{\mathbf{n}}$ by a single number, the best fit temperature $T(\hat{\mathbf{n}})$, and investigate how this temperature varies across the sky. In some instances, these two types of deviations are linked — we will return to this case when we discuss the so called SZ effect below.

What sort of spectral distortions might we theoretically expect? One recurring scenario in this thesis is that the IGM was reheated and reionized at some redshift $z < 10^3$. As long as the IGM temperatures are not too extreme, the thermal bremsstrahlung and double Compton processes will remain frozen out, so that the number of CBR photons will not change. However, the photons will Thomson scatter off free electrons, which will modify their energy slightly and change their direction of motion radically. As we will see in the following two sections, the former effect tends to cause spectral distortions whereas the latter effect tends to suppress spatial fluctuations.

2.5.3 Spectral distortions

In this subsection, we will assume that the universe is completely uniform and isotropic. With this assumption, the time-evolution of the photon spectrum $n_\gamma(\mathbf{r},\mathbf{k})$ in the presence of free electrons of temperature T_e is given by the Kompaneets equation (Kompaneets 1957)

$$\dot{n}_\gamma = \frac{kT_e}{m_e c^2}\sigma_t n_e c \frac{1}{x^2}\frac{\partial}{\partial x}\left[x^4\left(\frac{\partial n_\gamma}{\partial x} + n_\gamma + n_\gamma^2\right)\right],$$

where

$$x \equiv \frac{h\nu}{kT_e}.$$

This equation describes the effect of Compton scattering only, and is valid after about $z \sim 10^5$ when double Compton scattering and thermal Bremsstrahlung have frozen out so that the number of photons is conserved. Here $\mathbf{r}$ and $\mathbf{k}$ are taken to be the *comoving* quantities, so that the redshift term disappears from the equation and we simply have $\dot{n}_\gamma = 0$ if $n_e = 0$. Direct substitution shows that $\dot{n}_\gamma = 0$ also in the more interesting case where

$$n_\gamma = \frac{1}{e^{x+\mu} - 1}.$$

This means that if the photons start out with a Planck spectrum and if the electrons have the same temperature as the photons, then there will be no spectral distortions at all. Indeed, numerical integration of this nonlinear partial differential equation shows that for most physically reasonable initial spectra, the presence of the free electrons causes the photons to approach such a steady state. However, the resulting steady state typically has a "chemical potential" $\mu \neq 0$. Thus if some process such as the decay of a mysterious relic particle creates non-thermal photons in the early universe, the result will be a spectrum with $\mu \neq 0$. The FIRAS experiment has placed the strong limit $|\mu| < 3.3 \times 10^{-4}$ on the value of μ (Mather *et al.* 1994).

At the redshifts $z < 10^3$ that we probe in this thesis, this convergence towards a steady state is no longer effective, being basically frozen out. Suppose the free electrons have a temperature $T_e \neq T_\gamma$. Substituting this into the right-hand-side of the Kompaneets equation and using the fact that the resulting spectral distortions will be quite small, we obtain the first order approximation that the spectral distortion today will be simply (Zel'dovich & Sunyaev 1969)

$$\frac{\delta n_\gamma}{n_\gamma} \approx \left[\frac{x^2 e^x (e^x + 1)}{(e^x - 1)^2} - \frac{4xe^x}{e^x - 1}\right] y \approx \begin{cases} -2y & \text{for } x \ll 1, \\ x^2 y & \text{for } x \gg 1, \end{cases}$$

where

$$y \equiv \int \left(\frac{kT_e - kT_\gamma}{m_e c^2}\right) n_e \sigma_t c \, dt,$$

and the integral is to be taken from the early time when photon-creating processes froze out until today. Thus we see that to first order, the distortion will always have the same spectral profile, and that only the magnitude of the distortion depends on the thermal history of the electrons. This magnitude, referred to as the *Compton y-parameter*, thus summarizes the effect of free electrons by a single convenient number. Loosely speaking, the y-distortion indicates a relative excess of high-energy photons and a deficit of low-energy ones, which is caused by collisions in which hot electrons transfer energy to CBR photons.

There are two types of y-distortions of current observational interest. The first type is the average y-distortion over all of the sky, typically caused by very early reionization. This is the topic of Chapter 5. To date, no such y-distortion has been observed, and the upper limit of $y < 2.5 \times 10^{-5}$ places interesting constraints on the thermal history of the IGM. The second type is the localized y-distortion caused by hot plasma in rich clusters of galaxies. This is usually referred to as the Sunyaev-Zel'dovich effect, or the SZ effect for short, after the Russians who invented the y-parameter (Zel'dovich & Sunyaev 1969). The SZ-effect has indeed been observed in the direction of a number of large galaxy clusters — see for instance Birkinshaw & Hughes (1994). The experimental uncertainties are still large; around a factor of two. As soon as these shrink, the SZ-effect will become of great importance, since it provides an independent way of computing the distance to a galaxy cluster and thus can be used to determine the elusive Hubble constant. A recent review of the issues touched upon above is given by Hu & Silk (1993).

2.5.4 Spatial distortions

In this subsection, we will take the opposite approach: we will assume that there are no spectral distortions, so that the CBR that we observe from the direction $\hat{\mathbf{n}}$ in the sky has a Planck spectrum characterized by a temperature $T_\gamma(\hat{\mathbf{n}})$. It is the COBE observation of this function (Smoot *et al.* 1992) that has recently been plastered on front covers of magazines across the world and prompted various famous people to say silly things about holy grails and seeing the face of God. Our discussion of the ripples in $T_\gamma(\hat{\mathbf{n}})$, usually referred to simply as CBR fluctuations, will be organized as follows: first, we list some handy formulas for working with spherical harmonics. Then we review how the above-mentioned formulas for random fields are altered when dealing with a sphere instead of R^3. After this, we discuss how the CBR fluctuations are related to the fluctuations in the density field δ. Finally, we show how early reionization can suppress CBR fluctuations, which is relevant to Chapters 3, 4 and 5. A recent review of CBR fluctuations is given by White, Scott & Silk (1994).

Working with spherical harmonics

For functions that live on a sphere, the analogue of a Fourier expansion is an expansion in *spherical harmonics*. The spherical harmonics are defined as

$$Y_{lm}(\theta,\varphi) = \sqrt{\frac{2l+1}{4\pi}\frac{(l-m)!}{(l+m)!}}P_l^m(\cos\theta)e^{im\varphi},$$

where P_l^m are the associated Legendre functions, and l and m are integers such that $l \geq 0$ and $|m| \leq l$. They have the symmetry property that

$$Y_{l,-m} = (-1)^m Y_{lm}^*,$$

where $*$ denotes complex conjugation. These functions form a complete orthonormal set on the unit sphere. The orthonormality means that

$$\int Y_{lm}^*(\theta,\varphi)Y_{l'm'}^*(\theta,\varphi)d\Omega = \delta_{ll'}\delta_{mm'},$$

and the completeness means that we can expand any L^2 function Δ as

$$\Delta(\theta,\varphi) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l} a_{lm}Y_{lm}(\theta,\varphi),$$

where

$$a_{lm} \equiv \int Y_{lm}^*(\theta,\varphi)\Delta(\theta,\varphi)d\Omega.$$

Throughout this section, we use the differential solid angle notation

$$d\Omega = \sin\theta\; d\theta d\varphi.$$

Also, we will sometimes replace θ and φ by the unit vector

$$\hat{\mathbf{n}} = (\sin\theta\cos\varphi, \sin\theta\sin\varphi, \cos\theta)$$

and write things like

$$Y_{lm}(\theta,\varphi) = Y_{lm}(\hat{\mathbf{n}}).$$

With this notation, the so called addition theorem for spherical harmonics states that

$$\sum_{m=-l}^{l} Y_{lm}^*(\hat{\mathbf{n}})Y_{lm}(\hat{\mathbf{n}}') = \frac{2l+1}{4\pi}P_l(\hat{\mathbf{n}}\cdot\hat{\mathbf{n}}'),$$

where $P_l = P_l^0$ are the Legendre polynomials.

For the reader who wants more intuition about spherical harmonics, it is good to lump together all harmonics with the same l value. For l = 0, 1, 2, 3 and 4, these sets are referred to as the monopole, the dipole, the quadrupole, the octupole and the hexadecapole, respectively. Geometrically, when one expands a function $f(\hat{\mathbf{n}})$ in spherical harmonics, $l = 0$ picks up the constant part, $l = 1$ picks up the remaining linear part, $l = 2$ picks up the remaining quadratic part, $l = 3$ picks up the remaining cubic part, *etc.* In group theory jargon, the spherical harmonics corresponding to different l-values are the irreducible representations of the group of rotations of the sphere. This means that if one expands a rotated version of the same function f, the new spherical harmonic coefficients a'_{lm} will be some linear combination of the old ones a_{lm},

$$a'_{lm'} = \sum_{m=-l}^{l} D_{lm'm}\, a_{lm},$$

such that different l-values live separate lives and never mix. For instance, however one chooses to rotate a linear function ($l = 1$), it will always remain linear and never say pick up quadratic terms.

Random fields on the sphere

$T_\gamma(\hat{\mathbf{n}})$, the CBR temperature that we observe in the direction $\hat{\mathbf{n}}$ in the sky, is modeled as a random field. It is more convenient to work with the dimensionless version

$$\Delta(\hat{\mathbf{n}}) \equiv \frac{T_\gamma(\hat{\mathbf{n}})}{\langle T_\gamma(\hat{\mathbf{n}})\rangle} - 1,$$

which is often denoted $\Delta T/T$ in the literature. This field is related to the random density field δ that we discussed earlier on, and we will return to how they are relater further on. All the formulas and definitions we gave for δ have spherical analogues for Δ, as we will now see. Δ is usually assumed to be Gaussian, which follows from the random phase assumption

$$\langle a^*_{lm} a_{l'm'}\rangle = \delta_{ll'}\delta_{mm'} C_l,$$

where the coefficients C_l constitute the spherical version of the power spectrum $P(k)$. The *angular correlation function* is defined as

$$c(\theta) = \langle \Delta(\hat{\mathbf{n}})\Delta(\hat{\mathbf{n}}')\rangle,$$

where $\hat{\mathbf{n}}\cdot\hat{\mathbf{n}}' = \cos\theta$, and the right hand side is independent of the actual directions by the isotropy assumption. Just as the spatial correlation function was the Fourier transform of the spatial power spectrum, the angular correlation function is what might be called a "Legendre transform" of the angular power spectrum C_l. Using the addition theorem, one readily obtains

$$c(\theta) = \frac{1}{4\pi}\sum_{l=0}^{\infty}(2l+1)C_l P_l(\cos\theta).$$

And just as the r.m.s. fluctuations of the smoothed density field was given by integrating the power spectrum against a window function $\widehat{w}$, the r.m.s. temperature fluctuations seen by a CBR experiment with some angular selection function (beam pattern) is given by a weighted average of the coefficients C_l, the weights being a sort of "Legendre transform" of the beam pattern. Explicit examples of this are given in Appendix C, and a few selected window functions are plotted Figure 2.4.

The monopole and the dipole

The non-cosmologist reader may wonder why all sums over l in that chapter start with $l = 2$ rather than with $l = 0$. This is because we have no way of measuring the monopole coefficient c_0 or the dipole coefficient c_1.

If we knew the ensemble average $\langle T_\gamma \rangle$ (or could measure it by repeating the COBE experiment in many different horizon volumes throughout the universe and invoking ergodicity), then we could calculate our observed monopole as the difference between the average T_γ that we observe in our sky and $\langle T_\gamma \rangle$. But we can't... In other words, our CBR monopole is the difference between the average temperature in our sky and the average temperature of the entire universe. But since we use the former as our estimate of the latter, this gets us nowhere with the monopole. Since direct estimates of the age of the universe are still quite uncertain, but we know the CBR temperature to three decimal places, the CBR in a sense merely tells us what time it is, *i.e.* how much our universe had expanded by the time this civilization of ours turned up on the scene. Thus it seems highly unlikely that we will be able to use any kind of physics to predict $\langle T_\gamma \rangle$ accurately (to one part in 10^{-5} or so) in the foreseeable future.

The situation is similar with the dipole. If we knew that we were comoving observers, *i.e.* at rest with respect to the comoving FRW coordinates, then we could measure our local CBR dipole as the dipole we actually see in the sky. But the most accurate way we have to determine the comoving rest frame is to calculate the frame in which the CBR dipole vanishes. In other words, our CBR dipole is the difference between the average dipole in our sky and our velocity vector. Since we use the former as our estimate of the latter, this gets us nowhere with the dipole either. Thus the lowest multipole which gives us information about the power spectrum is $l = 2$, the quadrupole.

The relationship between δ and Δ

To date, the only really accurate way to compute the CBR power spectrum C_l is to integrate the Boltzmann equation numerically. Today it takes a few hours of CPU time on a high-end workstation to compute the first few thousand C_l-coefficients. The results for a CDM model and a BDM model are plotted in Figure 2.5, multiplied by $l(l+1)$. To give some intuition about the general shape of these functions, we will very briefly mention the main physical effects at work.

In the absence of reionization, the CBR photons that reach us today are carrying information about the electron distribution on the *last-scattering surface*, the spherical region around us where the photons last Thomson scattered off of an electron some some 10^{10} lightyears away. There are three mechanisms through which adiabatic density fluctuations δ on the last-scattering surface cause CBR anisotropies:

1. Fluctuations in the gravitational potential ϕ causing a gravitational redshift/blueshift

2. Bulk velocities on the last-scattering surface, causing Doppler shifts

3. Fluctuations in the photon density on the last-scattering surface

For adiabatic fluctuations, it turns out that the third effect cancels two thirds of the first effect. The net result in known as the Sachs-Wolfe effect, and for a matter power spectrum $P(k) \propto k^n$, the Sachs-Wolfe effect turns out to give

$$C_l \propto \frac{\Gamma\left(l+\frac{n-1}{2}\right)}{\Gamma\left(l+\frac{5-n}{2}\right)},$$

which for the case $n = 1$ reduces to simply

$$C_l \propto \frac{1}{l(l+1)}.$$

The Sachs-Wolfe effect is the dominant source of anisotropies on large angular scales $\theta \gg 1°$, corresponding to small l. This is why the $n = 1$ CDM curve in Figure 2.5 is flat for small l.

The funny-looking bumps in the CDM curve, known as "Doppler peaks" are due to a combination of effects 2 and 3. Finally, the last scattering surface has a finite thickness, which means that when we look in a given direction, we are in fact seeing a mixture of photons emanating from different points along that line of sight. This averaging of the fluctuations at many different points washes out fluctuations on very small angular scales, which is why all the curves approach zero for $l \gg 10^3$.

2.5.5 How reionization suppresses fluctuations

As mentioned in the introduction, reionization would affect the CBR in at least three ways:

1. The hot electrons would cause a y-distortion, as discussed above.

2. The fluctuations in the density and velocity of these hot electrons would generate new CBR anisotropies. For $\Omega = 1$, due to various cancellations, it turns out that this so called Vishniac effect becomes important only on very small angular scales. We will not discuss it further here — see Hu *et al.* 1994 for a thorough discussion.

3. Spatial fluctuations on angular scales below a few degrees could become suppressed, while fluctuations on larger scales such as those probed by COBE would remain fairly unaffected.

This third effect can be calculated numerically by integrating the linearized Boltzmann equation as mentioned above. Since hitting [RETURN] and waiting for hours does not greatly enhance ones intuitive understanding of the underlying physics, we will devote this section to deriving a simple analytic approximation which agrees fairly well with the numerical results for relatively late ($z < 200$) reionization. This is based on some ray tracing calculations made by the author. Examples of ray tracing work in the literature are Peebles (1987) and Anninos *et al.* (1991).

The ray tracing approach

Let $n_z(\mathbf{x}, \mathbf{q})$ denote the six-dimensional phase-space density of photons at redshift z that have comoving position $\mathbf{x}$ and wave vector $\mathbf{q} = q\hat{\mathbf{q}}$. If this function is known at our location

($\mathbf{x} = 0$) today ($z = 0$), it is straightforward to predict the outcome of any microwave background experiment. The time evolution of this function between early times and today is described by the Boltzmann equation for radiative transfer. Since this equation is linear in n and invariant under spatial translations, we can write

$$n_z(\mathbf{x},\mathbf{q}) = \int G(\mathbf{x}-\mathbf{x}',\mathbf{q},\mathbf{q}',z,z')n_{z'}(\mathbf{x}',\mathbf{q}')d^3x'd^3q' \tag{2.6}$$

for some integral kernel G. One way to interpret this G is as the probability density that a photon that has position $\mathbf{x}$ and momentum $\mathbf{q}$ at redshift z was at $\mathbf{x}'$ with momentum $\mathbf{q}'$ at redshift z'. Thus an alternative to integrating the Boltzman equation numerically is to evaluate this probability distribution by Monte Carlo simulations of trajectories of large numbers of photons. As we will see, it is sometimes possible to obtain good analytic approximations for G as well.

The geometrical smudging effect

Reionization of the intergalactic medium (IGM) causes CBR photons to Thomson scatter off of free electrons, which affects CBR fluctuations in the three ways listed above. The third of these effects, to which we are limiting our attention, is purely geometrical in nature, *i.e.* independent of the fluctuations in the electron temperature and velocity. Thus ignoring the frequency change that photons experience when they scatter, the magnitude q simply stays constant for all photons (there redshift is taken care of by the fact that we are using comoving position and momentum coordinates). Hence we can write equation (2.6) as

$$n_z(\mathbf{x},\hat{\mathbf{q}},q) = \int G(\mathbf{x}-\mathbf{x}',\hat{\mathbf{q}},\hat{\mathbf{q}}',z,z')n_{z'}\left(\mathbf{x}',\hat{\mathbf{q}}',q\right)d^3x'd\Omega', \tag{2.7}$$

where the angular integration is over all directions $\hat{\mathbf{q}}'$. Assuming Planck spectra and using the notation introduced in Section 2.5.4, this implies that

$$\Delta_z(\mathbf{x},\hat{\mathbf{q}}) = \int G(\mathbf{x}-\mathbf{x}',\hat{\mathbf{q}},\hat{\mathbf{q}}',z,z')\Delta_{z'}\left(\mathbf{x}',\hat{\mathbf{q}}'\right)d^3x'd\Omega'. \tag{2.8}$$

The isotropy approximation

A very useful approximation is that the radiation is isotropic early on, at the last scattering epoch $z = z_{rec}$. In this approximation, we can write

$$\Delta_{z_{rec}}(\mathbf{x},\hat{\mathbf{q}}) = \Delta_{rec}(\mathbf{x}),$$

since the left hand side is independent of $\hat{\mathbf{q}}$. For a thin last-scattering surface, it is easy to see that this is the case for the contribution from both the Sachs-Wolfe and the intrinsic density fluctuations, whereas the Doppler term will have a dipole anisotropy. With the isotropy approximation, we have

$$\Delta_z(\mathbf{x},\hat{\mathbf{q}}) = \int G(\mathbf{x}-\mathbf{x}',\hat{\mathbf{q}},z,z_{rec})\Delta_{rec}(\mathbf{x}')d^3x', \tag{2.9}$$

where we have defined the angularly averaged kernel

$$G(\Delta\mathbf{x},\hat{\mathbf{q}},z,z') \equiv \int G(\Delta\mathbf{x},\hat{\mathbf{q}},\hat{\mathbf{q}}',z,z')d\Omega'.$$

From here on, we will suppress z and z', taking $z = 0$ and $z' = z_{rec}$. Fourier transforming equation (2.9) with respect to $\mathbf{x}$ and using the convolution theorem leaves us with

$$\widehat{\Delta}_0(\mathbf{k}, \hat{\mathbf{q}}) = \widehat{G}(\mathbf{k}, \hat{\mathbf{q}})\widehat{\Delta}_{rec}(\mathbf{k}). \tag{2.10}$$

Thus there is no mode coupling at all, and for fixed $\hat{\mathbf{q}}$, the transfer function is simply the Fourier transform of the probability distribution $G(\Delta\mathbf{x}) \equiv G(\Delta\mathbf{x}, \hat{\mathbf{q}})$.

The transfer function $\widehat{G}$

Before we turn to the problem of explicitly evaluating $\widehat{G}$, we will briefly discuss its interpretation. Assuming that Δ is a Gaussian random field, it is easy to show that

$$\langle\widehat{\Delta}_0(\mathbf{k}, \hat{\mathbf{q}})\widehat{\Delta}_0(\mathbf{k}', \hat{\mathbf{q}})\rangle = |\widehat{G}(\mathbf{k}, \hat{\mathbf{q}})|^2\langle\widehat{\Delta}(\mathbf{k})_{rec}\widehat{\Delta}_{rec}(\mathbf{k}')\rangle,$$

i.e. that the quantity $|\widehat{G}(\mathbf{k}, \hat{\mathbf{q}})|$ plays the role of a transfer function. Thus the radiation power spectrum on scales $\lambda = 2\pi/k$ is suppressed by a factor $|\widehat{G}(\mathbf{k}, \hat{\mathbf{q}})|^2$. Let us make a few elementary observations that are valid for an arbitrary probability distribution G.

Observation (I): $|\widehat{G}(\mathbf{k})| \leq 1$, with equality if $\mathbf{k} = 0$ or $G(\Delta\mathbf{x}) = \delta(\Delta\mathbf{x} - \Delta\mathbf{x}_0)$.

Observation (II): $\widehat{G}(\mathbf{k}) \to 0$ as $|\mathbf{k}| \to \infty$ if G is an integrable function (as opposed to say a tempered distribution like $\delta(\Delta\mathbf{x} - \Delta\mathbf{x}_0)$).

Observation (III): If we define $\mathbf{a}$ to be the mean and S to be the covariance matrix of the probability distribution G, then

$$\widehat{G}(\mathbf{k}) = 1 - ia_m k_m - \frac{1}{2}(a_m a_n + S_{mn})k_m k_n + O(|\mathbf{k}|^3),$$

so

$$|\widehat{G}(\mathbf{k})|^2 = 1 - S_{mn}k_m k_n + O(|\mathbf{k}|^3). \tag{2.11}$$

(Repeated indices are to be summed over, from 1 to 3.)

Observation (I) tells us that Thomson scattering causes a low-pass filtering of the Fourier modes, *i.e.* leaves very long wavelengthes unaffected. $G(\Delta\mathbf{x}) = \delta(\Delta\mathbf{x} - \Delta\mathbf{x}_0)$ corresponds to no reionization, in which case there is no smudging at all and no modes are suppressed. Observation (II), which is known as Riemann-Lebesgue's Lemma, tells us that unless there is a finite probability for no scattering at all, very high frequency modes get almost entirely damped out.
Observation (III) gives us a Taylor expansion of $\widehat{G}(\mathbf{k})$ around the origin, which will prove useful below.

All quantities of physical interest (C_l, $C(\theta)$, *etc.*) can be computed from the initial power spectrum at z_{rec} once the transfer function $\widehat{G}(\mathbf{k}, \hat{\mathbf{q}})$ is known, and explicit formulas for this are given in most standard texts. A crude rule of thumb simply identifies the comoving length scale λ with the angle that this distance subtends on the last scattering surface, a spherical region whose radius is our roughly our horizon radius $a_0\tau_0 \approx 6000h^{-1}$Mpc. Another rule of thumb is to identify a multipole moment ℓ with the angular scale $\theta \approx 180°/l$. Thus for small angles, we have the rough correspondence

$$\lambda \propto \theta \propto \frac{1}{l} \propto \frac{1}{k},$$

where the proportionality constants are given by the correspondence

$$\theta \approx 1° \quad \sim \quad l \approx 200 \quad \sim \quad \lambda \approx 100h^{-1}\mathrm{Mpc}.$$

Let us define the *smudging scale*

$$\lambda_c \equiv (\det S)^{1/6},$$

i.e. as the geometric mean of the three eigenvalues of $S^{1/2}$. For realistic scenarios, the smudging tends to be fairly isotropic. Thus all three eigenvalues of the covariance matrix S tend to be of the same order of magnitude, and the Taylor expansion (2.11) gives

$$|\widehat{G}(\mathbf{k})|^2 \approx 1 - (\lambda_c k)^2 \tag{2.12}$$

for $k \ll \lambda_c^{-1}$. When we detect a CBR photon arriving from some direction $\hat{\mathbf{q}}$ in the sky and ask where it was at $z = z_{rec}$, the smudging scale λ_c is roughly the standard deviation of the answer (see Figure 2.7).

Approximating the transfer function $\widehat{G}$

For fairly late (say $z < 200$) reionization, there is a substantial probability that a given CBR photon manages to dodge all the reionized electrons without being scattered a single time. Writing this probability as $e^{-\tau_t}$, where the quantity τ_t is called the *optical depth* and will recur in subsequent chapters, the probability that a photon is scattered n times will be $e^{-\tau_t}\frac{\tau_t^n}{n!}$, a Poisson distribution. (τ_t is not to be confused with the conformal time τ.) Let us expand the propagator G in a type of Born expansion

$$G = \sum_{n=0}^{\infty} e^{-\tau_t}\frac{\tau_t^n}{n!}G_n, \tag{2.13}$$

where G_n is the probability distribution corresponding to the case where the photon scatters exactly n times. Omitting the trivial dependence on $\hat{\mathbf{q}}$ and $\hat{\mathbf{q}}'$, we have simply

$$G_0(\Delta\mathbf{x}) = \delta(\Delta\mathbf{x} - \Delta\mathbf{x}_0),$$

where $\Delta\mathbf{x}_0$ is some constant, since if the photon never scattered, we know exactly where it came from. G_1 can be readily calculated analytically, which we will do in the next section, after which the higher order terms can be calculated by repeated integrations if desired. A nice feature is that for late reionization, τ_t is often so small that a good approximation is obtained by dropping all but a few of the first terms in equation (2.13). For instance, reionization around $z = 50$ with $\Omega = 1$ and $h\Omega_b = 0.03$ gives $e^{-\tau_t} \approx 0.70$, $e^{-\tau_t}\tau_t \approx 0.25$ and $e^{-\tau_t}\frac{\tau_t^2}{2} \approx 0.044$, so if we keep only the G_0 and G_1 terms, we will be off by only 5%, and if we include the G_2 term as well, we are off by less than 1%. Physically, this simply means that if it is not that likely that a photon scattered even a single time, then it is quite unlikely that it scattered more than a few times, so that we can safely neglect that possibility. If one nonetheless wants to include all terms, it is readily done by Monte Carlo simulation that follow individual photons as they propagate upwards in Figure 2.1, occasionally changing direction as they Thomson scatter off of free electrons. Some examples of this are shown in Figures 2.6 and 2.7. Let us write the infinite sum (2.13) as

$$G = e^{-\tau_t} G_0 + \left(1 - e^{-\tau_t}\right) G_{1+},$$

where G_{1+} contains the contributions from all terms with $n \geq 1$. We Fourier transform this with as before:

$$\widehat{G}(\mathbf{k}, \hat{\mathbf{q}}) = e^{-\tau_c} + \left(1 - e^{-\tau_c}\right) \widehat{G}_{1+}(\mathbf{k}, \hat{\mathbf{q}})$$

Since $\widehat{G}_{1+}$ contains no delta-functions, its Fourier transform approaches zero as $|\mathbf{k}| \to \infty$ according to Observation I. Thus

$$\widehat{G}(\mathbf{k}, \hat{\mathbf{q}}) \to e^{-\tau_c} \quad \text{as} \quad k \to \infty.$$

Combining this with the estimate (2.11), we thus know how the transfer function behaves in both limits: the approximation

$$|\widehat{G}(\mathbf{k}, \hat{\mathbf{q}})| \approx \left(1 - e^{-\tau_t}\right) e^{-\frac{1}{2}\mathbf{k}^T S \mathbf{k}} + e^{-\tau_t} \tag{2.14}$$

becomes exact both as $k \to 0$ and as $k \to \infty$. Using equation (2.12) yields the isotropic approximation

$$|\widehat{G}(\mathbf{k}, \hat{\mathbf{q}})| \approx \left(1 - e^{-\tau_t}\right) e^{-(\lambda_c k)^2/2} + e^{-\tau_t}. \tag{2.15}$$

This function is plotted in Figure 2.8 for a few reionization scenarios. The values of λ_c used in this plot where computed through Monte Carlo simulation in conjunction with Wayne Hu.

The backward light cone

We will conclude our discussion of how reionization suppresses fluctuations by computing the function G_1 analytically. Let us assume that $\Omega = 1$. To simplify the calculations that follow, we will use comoving coordinates $\mathbf{x}$ and conformal time τ, so that all light rays (null geodesics) make 45 degree angles with the τ axis in $(\tau, \mathbf{x})$-space, just as in Penrose diagrams. Let us choose the scale factor a_0 such that $\tau = 1$ corresponds to today, *i.e.* so that

$$\tau = \frac{1}{\sqrt{1+z}}.$$

The situation is illustrated in Figure 2.1, where we have suppressed one of the three spatial coordinates for the sake of the plot. We are at the apex of the cone, our *backward light cone*, and in the absence of reionization, the CBR photons that reach us today have been moving on the surface of this cone ever since recombination.

Where where the photons at recombination?

If we detect a CBR photon arriving from a given direction in the sky, where was is at a redshift z? The presence of free elections makes it possible for a photon to change direction a number of times by inelastic Thomson scattering. We neglect other ways by which photons can change direction, such as gravitational lensing.

To avoid double use of the letter z, we will write comoving coordinates as $\mathbf{x}=(u,v,w)$. Suppose a photon arrives from the w-direction to the point $\mathbf{x}=0$ at some conformal time τ. Where was it at the earlier time $\tau_1=\tau-\Delta\tau$? If it has not been scattered during the interval $\Delta\tau$, it has followed a null geodesic, and its position at τ_1 is given by

$$\mathbf{x}^{(0)}(\tau_1)=(0,0,\Delta\tau). \tag{2.16}$$

If it has been scattered exactly once during the interval, say at the conformal time $\tau_1+\tau_s$, and previously came from the direction (θ,ϕ) in spherical coordinates, then

$$\mathbf{x}^{(1)}(\tau_1)=(u,v,w)=\left[\tau_s\sin\theta\cos\phi,\tau_s\sin\theta\sin\phi,\tau_s\cos\theta+(\Delta\tau-\tau_s)\right]. \tag{2.17}$$

For $|\mathbf{x}|\leq\Delta\tau$, this is readily inverted, yielding

$$\begin{cases}\tau_s &= \frac{u^2+v^2+(\Delta\tau-w)^2}{2(\Delta\tau-w)}\\ \theta &= 2\arctan\frac{\Delta\tau-w}{\sqrt{u^2+v^2}}\\ \phi &= \arctan\frac{v}{u}\end{cases} \tag{2.18}$$

To answer the question "where did the photon come from if it scattered exactly once", let us take τ_s θ and ϕ to be independent random variables with probability distributions f_τ, f_θ and f_ϕ, and calculate the probability distribution for the random variable $\mathbf{x}(\tau_1)$ as defined by equation (2.17). This probability distribution is

$$G_1(\mathbf{x})=\begin{cases}f_\tau\left[\tau_s(\mathbf{x})\right]f_\theta\left[\theta(\mathbf{x})\right]f_\phi\left[\phi(\mathbf{x})\right]\frac{\partial(\tau_s,\theta,\phi)}{\partial(u,v,w)} & \text{if } |\mathbf{x}|\leq\Delta\tau,\\ 0 & \text{if } |\mathbf{x}|>\Delta\tau,\end{cases} \tag{2.19}$$

where the volume element is

$$\frac{\partial(\tau_s,\theta,\phi)}{\partial(u,v,w)}\equiv\left|\begin{array}{ccc}\frac{\partial\tau_s}{\partial u}&\frac{\partial\tau_s}{\partial v}&\frac{\partial\tau_s}{\partial w}\\ \frac{\partial\theta}{\partial u}&\frac{\partial\theta}{\partial v}&\frac{\partial\theta}{\partial w}\\ \frac{\partial\phi}{\partial u}&\frac{\partial\phi}{\partial v}&\frac{\partial\phi}{\partial w}\end{array}\right|=\frac{1}{\sqrt{u^2+v^2}(\Delta\tau-w)}.$$

For Thomson scattering, the angular dependence is $\frac{1}{\sigma}\frac{d\sigma}{d\Omega}=\frac{3}{16\pi}\left(1+\cos^2\theta\right)$, so

$$\begin{cases}f_\theta(\theta) &= \int_0^{2\pi}\frac{3}{16\pi}\left(1+\cos^2\theta\right)\frac{d\Omega}{d\theta d\phi}d\phi=\frac{3}{8}\left(1+\cos^2\theta\right)\sin\theta,\\ f_\phi(\phi) &= \frac{1}{2\pi}.\end{cases}$$

The function f_τ depends on the ionization history of the universe, so we will leave it arbitrary for the time being. Figure 2.1 shows a situation where the photon was scattered at $z=10$: we then know that it must have arrived on the smaller backward light cone emanating from the scattering event. This corresponds to the probability distribution $f_\tau(\tau)$ being a delta function. If we know that the photon scattered sometime around $z\approx 10$, and exactly once, but do not know exactly when, we obtain a probability distribution $G_1(\mathbf{x})$ that is smeared out around the perimeter of this $z=10$ backward light cone. For more physically realistic cases, where scattering occurs predominantly at higher redshifts, the resulting probability distribution tends to be unimodal.

2.6 Cosmological Chemistry

Almost all of the following chapters will involve the chemistry of the IGM, the intergalactic medium, and frequent references will be made to constraints from nucleosynthesis. The purpose of this section is to convince the reader that it is quite straightforward to understand chemical equations, even if one has never been anywhere near a test tube.

2.6.1 Evolution of the IGM

As an example, let us take the evolution of the ionization state of the IGM in the simplest case, neglecting helium and ionizing photons. Let x denote the ionization fraction and n the total number density of protons, free and bound. Then the densities of hydrogen atoms, free electrons and free protons are $n_H = (1-x)n$, $n_e = xn$ and $n_p = xn$, respectively. Thermal ionization occurs through the reaction

$$e^- + H \mapsto e^- + e^- + p^+,$$

i.e. the electron is a catalyst that collides with the hydrogen atom and uses its kinetic energy to ionize it. Thus the rate per unit volume for this reaction is proportional to $n_e n_H = n^2 x(1-x)$. Similarly, the rate for the recombination reaction

$$e^- + p^+ \mapsto H + \text{photon}$$

is proportional to $n_e n_p = n^2 x^2$. Thus the time-evolution of x is

$$\dot{x} = n[Ax(1-x) - Bx^2], \tag{2.20}$$

where the proportionality constants A and B are functions of temperature (since the particles come near each other more often if they are moving faster, and since the reaction cross sections depend on the kinetic energies of the particles). The cosmologist generally takes such functions from the atomic physics literature without worrying to much about how they were calculated. In addition to the equation specifying $\dot{x}$, one also needs an equation for $\dot{T}$, to be able to compute the evolution of temperature. Such an equation is quite straightforward to interpret, the various terms simply corresponding to the various cooling and heating effects that one wishes to take into account.

Before integrating the above-mentioned equations numerically, it is often convenient to change the independent variable from time to redshift. Then the IGM density is simply given by $n = n_0(1+z)^3$, n_0 denoting the density today. Using equation (2.1) for dz/dt, equation (2.20) becomes

$$-\frac{dx}{dz} = H_0 n_0 \frac{1+z}{\sqrt{1+\Omega_0 z}}[Ax(1-x) - Bx^2].$$

2.6.2 Nucleosynthesis

For a more complicated problem like nucleosynthesis [11], the formation of light elements during the first three minutes after the Big Bang, the mathematical structure of the problem

[11] This calculation would usually be referred to as nuclear physics rather than chemistry, but the resulting equations are of the same type — the only difference being that nucleons (protons and neutrons) rather than atoms are the fundamental entities.

is the same, merely more complicated. Instead of one single fraction x to keep track of, there will be a whole bunch of fractions x_i, $i = 1, 2, ..., n$: the fraction of protons in hydrogen, the fraction of protons in helium-3, the fraction of protons in helium-4, the fraction of protons in lithium, *etc.* The time derivative of each one of them is given by an equation like the one above, where the left hand side is $\dot{x}_i$ and the right hand side contains one term for each process that destroys or creates species i. The resulting system of n coupled first-order ordinary differential equations is fed into a computer and solved numerically. The numerical part is somewhat complicated by the fact that chemical equations tend to be stiff, but there is a plethora of standard packages that handle this sort of problems, for instance in the NAG library.

When this calculation is carried out — see Malaney & Mathews (1993) for a recent review, results are obtained that are in fairly good agreement with the primordial abundances of helium-4, helium-3, deuterium and lithium that are inferred from observational data. Pioneered by George Gamow and others in the forties (Alpher, Bethe & Gamow 1948[12]), nucleosynthesis predictions are one of the most striking successes of the Big Bang model. Apart from being a powerful weapon against rival cosmological theories, nucleosynthesis calculations also place an important constraint on the standard cosmological model, since the results turn out to depend strongly on the ratio between the densities of baryons and microwave background photons. This translates into a strong dependence on $h^2\Omega_b$. The most recent calculations to date (Smith *et al.* 1993; Walker *et al.* 1991) give the 95% confidence interval

$$0.010 < h^2\Omega_b < 0.015.$$

(Levels of confidence must be taken with a grain of salt in cosmology, where systematic errors often dominate.) This interval keeps shifting around and getting narrower as observational abundance data gets better and the the knowledge of the nuclear cross sections that go into the theoretical calculation improves.

2.7 Astronomy Jargon

Since the scales involved in cosmology are many orders of magnitude larger than those usually encountered in physics, cosmologist tend to use larger units of measurement. A convenient unit of time is the Giga-year:

$$1\text{Gyr} = 10^9 \text{ years} \approx 3.15 \times 10^{16} s \approx 10^{16}\pi \ s.$$

An *astronomical unit*, or AU for short, is the average distance between the earth and the sun.

$$1\text{AU} \approx 1.4960 \times 10^{13}\text{cm}.$$

For historical reasons, related to the parallax triangulation method by which the distance to nearby stars is calculated, the standard astronomical distance unit is the *parsec*. This is

[12] Gamow added his friend Hans Bethe to the author list of this so called $\alpha\beta\gamma$ paper as a joke...

the distance at which an AU makes an angle of one arcsecond (1/3600 of a degree) in the sky. Thus

$$1\text{pc} = \frac{360 \times 3600}{2\pi}\text{AU} \approx 3.086 \times 10^{18}\text{cm} \approx 3.2615 \text{ light-years.}$$

This is still a tiny distance by cosmology standards. The standard length unit in cosmology is the *megaparsec*:

$$1\text{Mpc} \approx 3.086 \times 10^{24}\text{cm}.$$

The standard astrophysical unit of mass is *solar mass*, denoted $M_\odot$. The mass of the sun is

$$M_\odot \approx 1.989 \times 10^{33}\text{g}.$$

Apart from involving many powers of ten, some cosmological parameters are quite uncertain. It is convenient to write such parameters as the product of a large round number that is in the right ballpark and a small dimensionless "fudge factor". For instance, the dimensionless quantity h is ubiquitous in the cosmology literature. This is because the Hubble constant is factored as

$$H = \left(\frac{100\,\text{km/s}}{\text{Mpc}}\right) h.$$

Whereas Hubble originally estimated that $h \approx 5$, most cosmologists today assume that $0.5 < h < 0.8$. The jury is still out. If some form of matter x has a uniform density ρ_x, then this density is usually factored as

$$\rho_x = \rho_{crit}\Omega_x,$$

where

$$\rho_{crit} \equiv \frac{3H^2}{8\pi G} \approx 1.9 \times 10^{-29} h^2 \frac{\text{g}}{\text{cm}^3}$$

is the critical density mentioned above. Thus Ω_b is a measure of the baryon density, Ω_{igm} is a measure of the IGM density, *etc.*

The only other unexplained astronomy jargon that I have been able to spot in the subsequent chapters is some reference to different types of stars. For historical reasons (ever heard that excuse before...?), stars are classified in categories called O, B, A, F, G, K, and M, where O-stars are the hottest and M-stars the coolest. Within each category, there are ten sub-categories specified by numbers, where for instance an O3 star is hotter than an O9 star.

2.8 An Abstract for Non-Cosmologists

So in plain English, what is this thesis all about?

It is now widely accepted that something heated and ionized the hydrogen that fills intergalactic space, but we don't know quite when this happened.

If this is to have had interesting effects on the microwave background fluctuations, then it must have occurred quite long ago, at redshifts greater than 50 or so. This is the topic of Chapters 3 and 4.

Some theories predict that the hydrogen has indeed been ionized pretty much all along. In Chapter 5 we show to what extent such scenarios are ruled out by recent measurements of the microwave background spectrum.

To affect the microwave background, 100% ionization is not required — some reasonable fraction like 30% is quite sufficient, providing that the ionization happens early enough. By studying the spectra from quasars, however, one gets more extreme constraints: in "recent" times, for redshifts less than about five, it appears that at least 99.999% of the hydrogen atoms must have been ionized. It is still not entirely clear whether photoionization alone is able to produce the high temperatures required to achieve this. In Chapter 6 we provide an alternative model for how this may have happened.

Finally, there is increasing evidence that the primordial power spectrum may be a more complicated function than a simple power law, an assumption that has been used in much of the literature, including Chapters 3, 4 and 6. Motivated by this evidence, the final chapter is devoted to developing a formalism that allows cosmological constraints to be placed that are completely independent of the shape of power spectrum. Some supplementary material for Chapters 3 and 7 is given in the appendices.

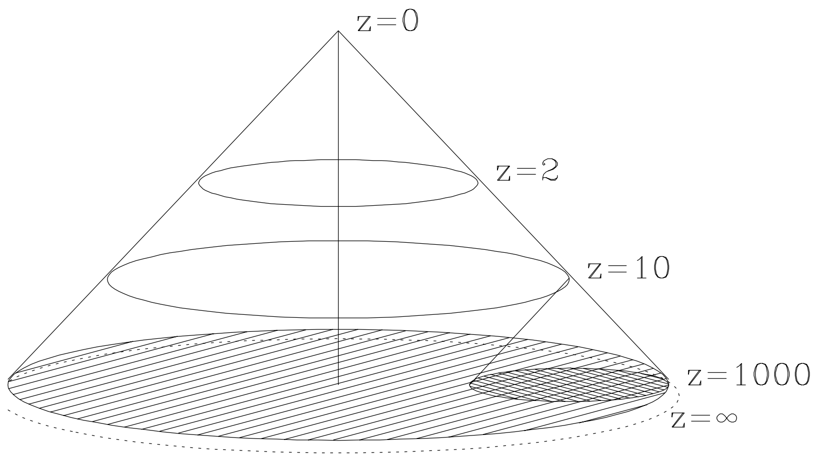


Figure 2.1: Our backward light cone.

Our backward light cone in a flat $\Omega = 1$ universe is shown in comoving spatial coordinates, with conformal time on the vertical axis (one of the three spatial dimensions has been suppressed). As can be seen, in these coordinates, light rays propagate in 45° lines just as in Euclidean space. The horizontal circles are labeled with their corresponding redshifts, the dotted circle corresponding to the Big Bang. If we (at the apex of the cone) detect a CBR photon arriving from the right in the picture, it could have been anywhere in the shaded region at $z = 10^3$, the recombination epoch. If we know that it has not been scattered since $z = 10$, it must have come from the double-hatched region.

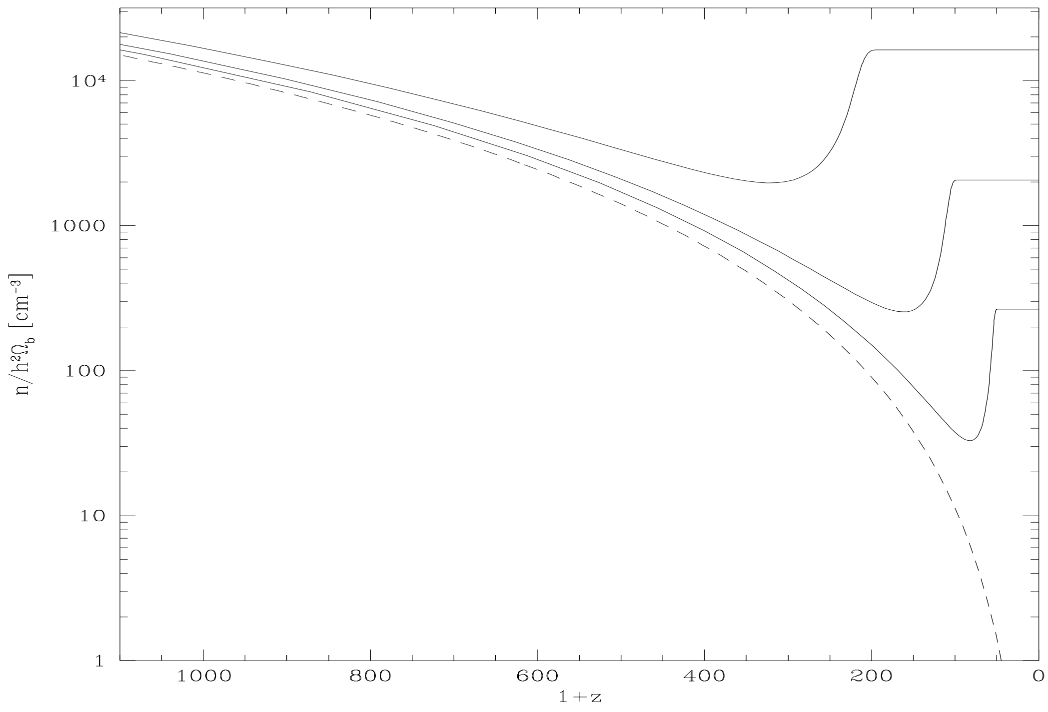


Figure 2.2: Evolution of a top hat overdensity.

The time-evolution of the density in a spherical ("top hat") overdense region in a flat universe is plotted as a function of redshift z. From top to bottom, the three solid curves correspond to virialization redshifts z_{vir} of 200, 100 and 50, respectively. The dashed curve gives the density of the IGM. The number density is given in nucleons per unit volume, *i.e.* $n = \rho/m_p$.

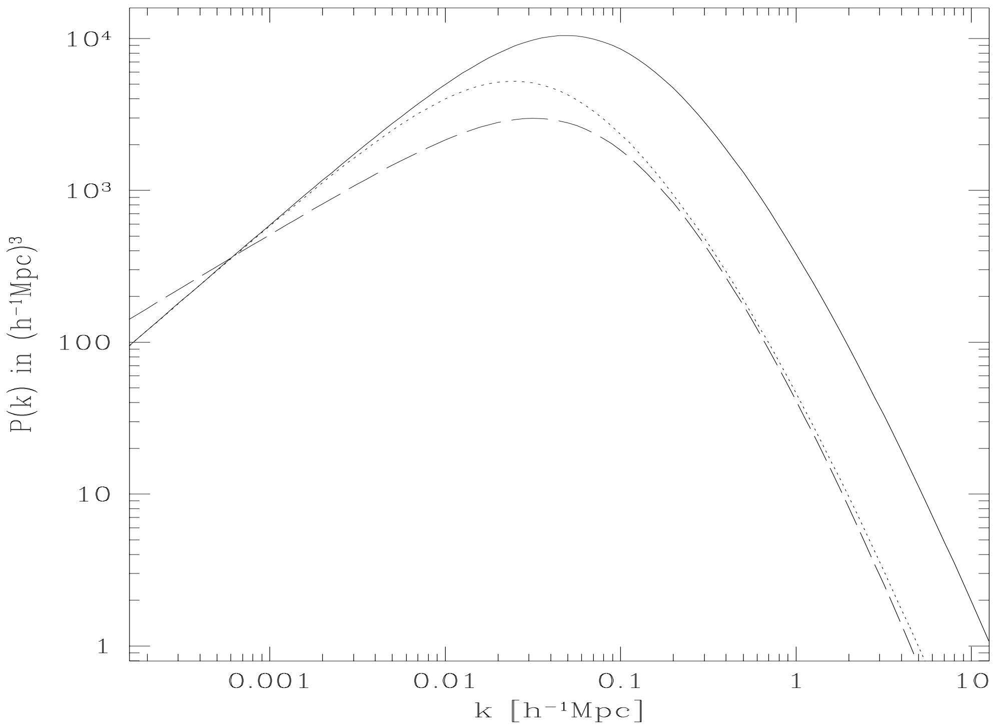


Figure 2.3: Assorted power spectra.

The power spectrum $P(k)$ is plotted for a CDM model and two variations on the theme. The solid curve is a standard CDM power spectrum with $h = 0.5$, $\Omega_0 = 1$, and $n = 1$ (Bond & Efstathiou 1984). The dotted curve has a lower "shape parameter" $h\Omega_0 = 0.25$, which can be attained in for instance the MDM and τCDM scenarios discussed in subsequent chapters, and corresponds to sliding the CDM curve to the left. The dashed curve has a lower spectral index $n = 0.7$, the so called tilted model, which approximately corresponds to rotating the CDM curve clockwise. All three power spectra have been normalized to COBE.

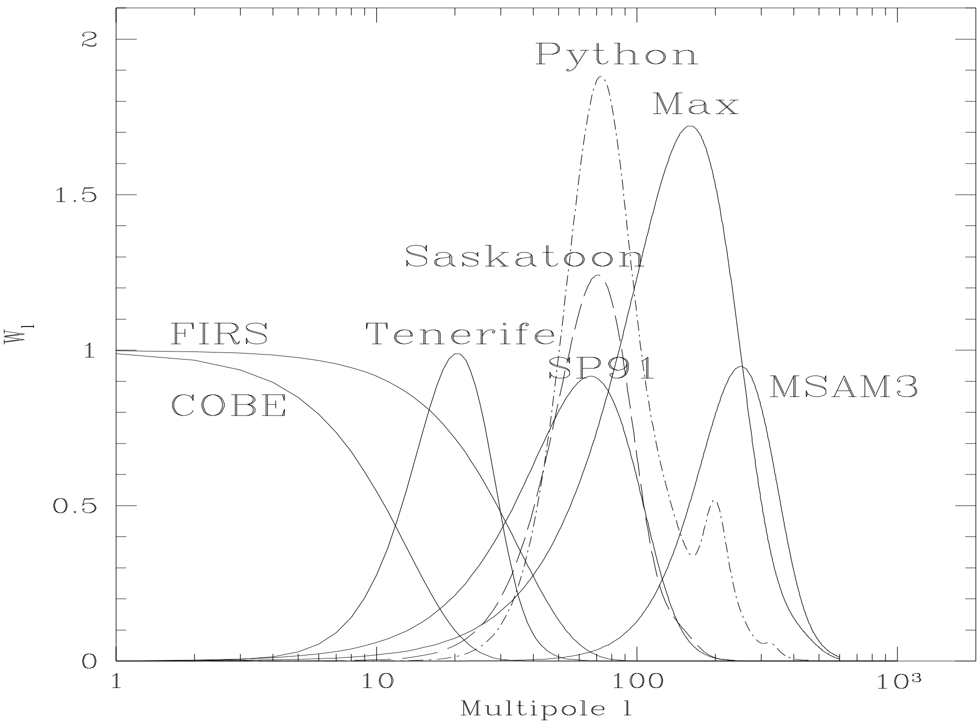


Figure 2.4: Assorted angular windows functions.

The angular window functions W_l are plotted for a number of recent CBR experiments. The data is courtesy of White, Scott & Silk (1994).

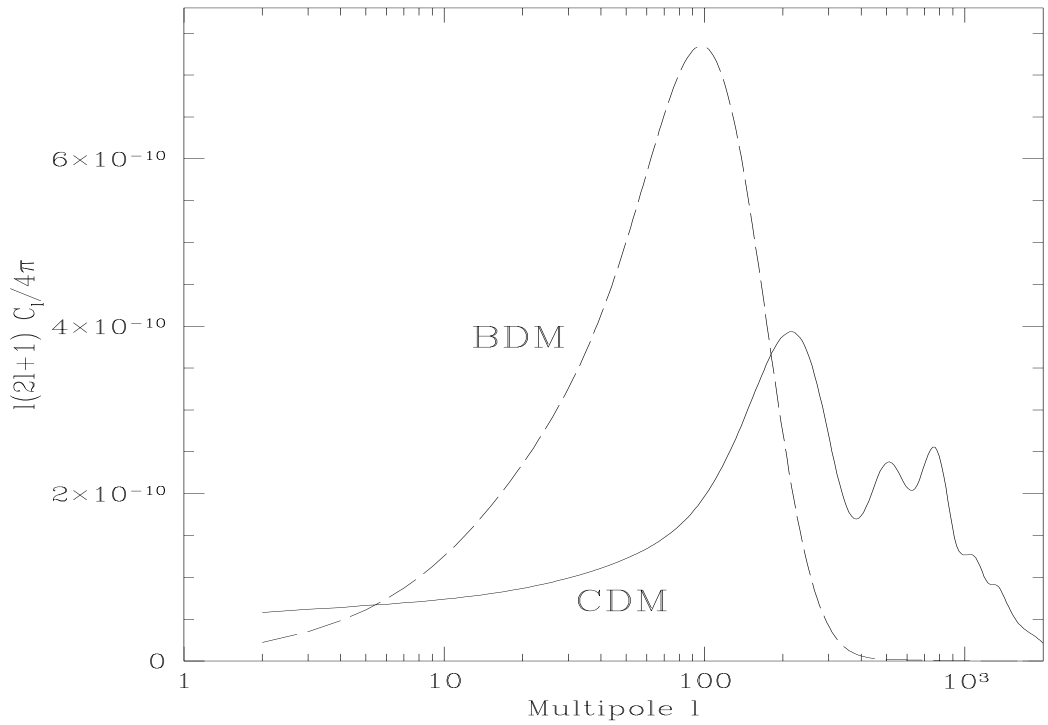


Figure 2.5: Assorted angular power spectra.

The angular power spectrum C_l is plotted for a CDM and a BDM models. The CDM model has $\Omega_0 = 1$, $\Omega_b = 0.06$, $h = 0.5$ and $n = 1$. The BDM model has $\Omega_0 = \Omega_b = 0.15$, $h = 0.8$ and $n = -0.5$. Both are normalized to COBE, *i.e.* give the same value when integrated against the window function for the COBE DMR 10° pixel variance. The models are courtesy of Naoshi Sugiyama and Wayne Hu.

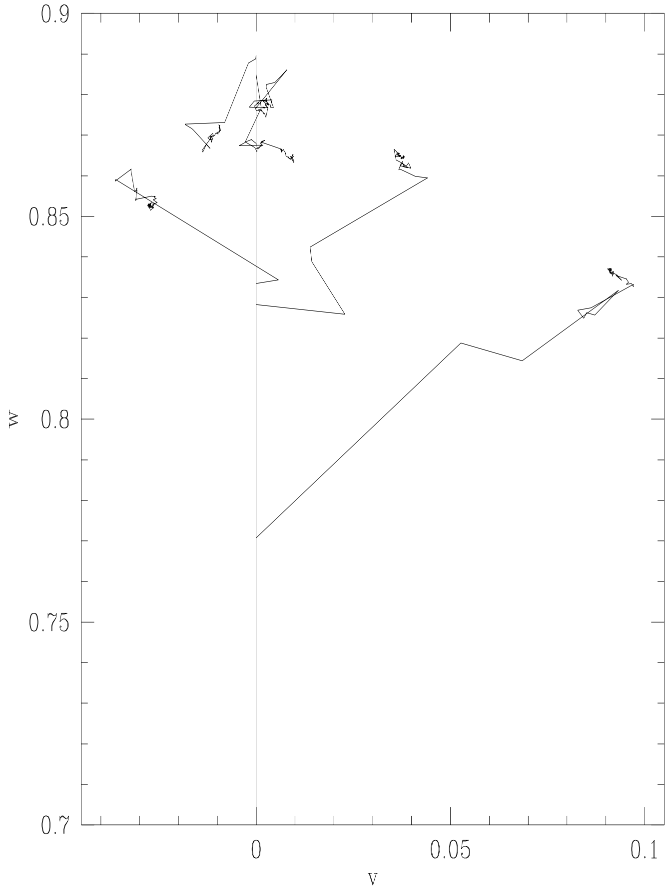


Figure 2.6: Monte Carlo photons, top view.

The trajectories of seven photons in a Monte Carlo simulation of a fully ionized universe are shown in comoving coordinates (u, v, w), with the u-coordinate perpendicular to the page. (In terms of Figure 2.1, this is a view from "above".) The question being asked is where photons arriving from the w-direction were in the past, so one can just as well model the photons as emanating from the apex of the light cone and traveling backwards in time. With this terminology, they all enter this figure from below, and part ways as they Thomson scatter. Notice how their mean free path gets shorter and shorter, which is due to the increase in electron density at higher redshifts.

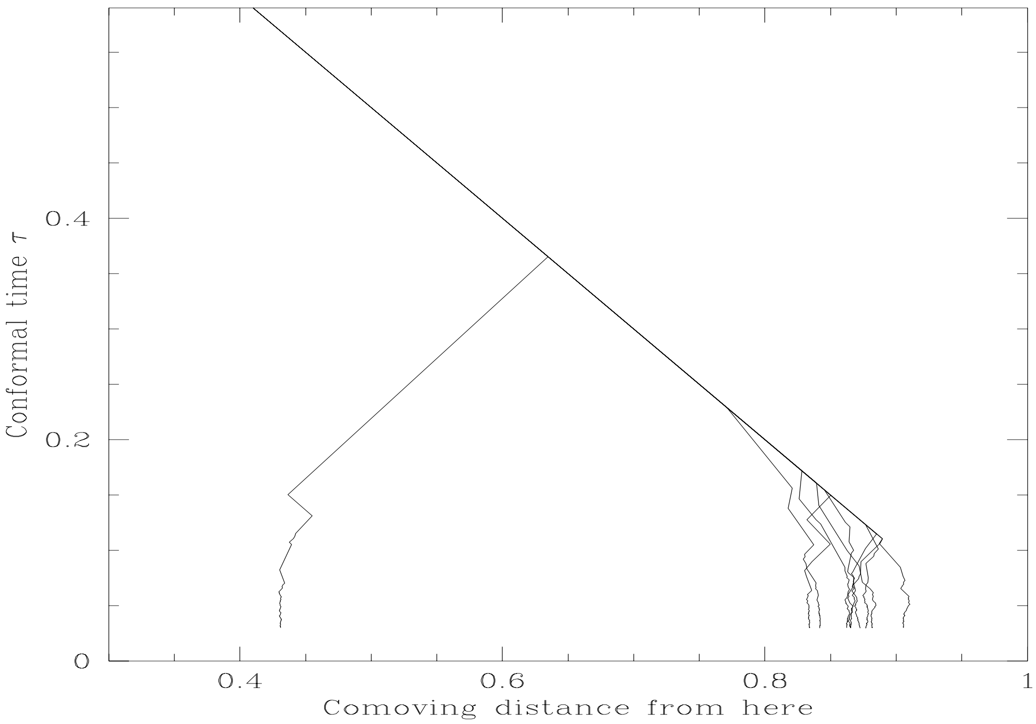


Figure 2.7: Monte Carlo photons, side view.

The trajectories of ten photons in the Monte Carlo simulation described in Figure 2.6 are plotted "from the side" with reference to the light cone in Figure 2.1. The horizontal axis is $(u^2 + v^2 + w^2)^{1/2}$, the comoving distance from "here" (the apex of the light cone). The curves all end at $\tau \approx 0.3$, corresponding to $z \approx 10^3$, *i.e.* standard recombination. The horizontal spread of these endpoints roughly corresponds to the smudging scale λ_c.

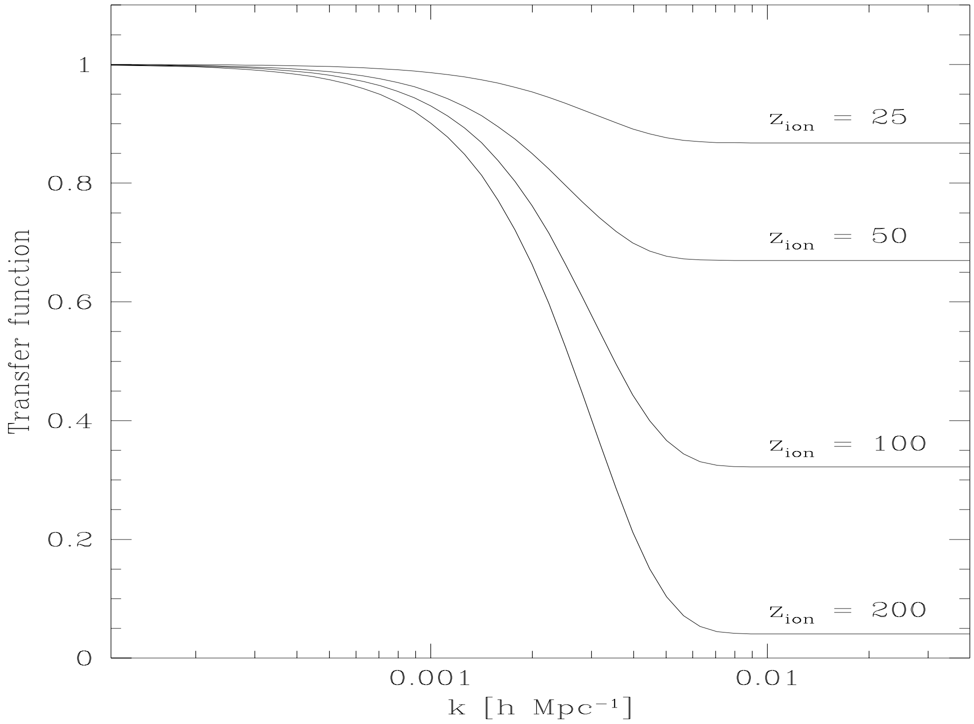


Figure 2.8: The transfer function $|\widehat{G}|$.

The transfer function $|\widehat{G}|$, *i.e.* the factor by which the r.m.s. CBR fluctuations are suppressed by reionization, is plotted as a function of comoving wavenumber for scenarios where the universe becomes reionized at some redshift z_{ion}. As $k \to \infty$, the suppression factor simply approaches $e^{-\tau_t}$. For these scenarios, $h\Omega_{igm} = 0.03$, so $\tau_t \approx (z_{ion}/92)^{3/2}$. The cutoff scales λ_c where determined by Monte Carlo simulation, and turn out to be depend only quite weakly on the ionization redshift z_{ion}. In all cases, degree-scale anisotropies are seen to be suppressed whereas fluctuations on the scale probed by COBE are almost unaffected.

Chapter 3

Early Reionization and CBR Fluctuations

In this chapter, early photoionization of the intergalactic medium is discussed in a nearly model-independent way, in order to investigate whether early structures corresponding to rare Gaussian peaks in a CDM model can photoionize the intergalactic medium sufficiently early to appreciably smooth out the microwave background fluctuations. We conclude that this is indeed possible for a broad range of CDM normalizations and is almost inevitable for unbiased CDM, provided that the bulk of these early structures are quite small, no more massive than about $10^8 M_\odot$. Typical parameter values predict that reionization occurs around $z = 50$, thereby suppressing fluctuations on degree scales while leaving the larger angular scales probed by COBE relatively unaffected. However, for non-standard CDM, incorporating mixed dark matter, vacuum density or a tilted primordial power spectrum, early reionization plays no significant role.

3.1 Introduction

The first quantitative predictions of cosmic microwave background anisotropies in cold dark matter (CDM)-dominated cosmological models recognized that reionization by rare, early-forming objects could play a role in suppressing temperature fluctuations on small angular scales (Bond & Efstathiou 1984; Vittorio & Silk 1984). Now that the COBE DMR experiment has detected fluctuations on large angular scales (Smoot *et al.* 1992) at a level (within a factor of two) comparable to that predicted by CDM models, it is especially relevant to examine whether reionization can affect the degree scale searches that are currently underway.

Cold dark matter models are generally characterized by a late epoch of galaxy formation. However, the smallest and oldest objects first go nonlinear at relatively large redshift. In this chapter we investigate, for a wide range of CDM normalizations, power spectra and efficiency parameters, whether reionization associated with energy injection by early forming dwarf galaxies can reionize the universe sufficiently early to smooth out primordial CBR temperature fluctuations.

Although we go into some detail in Appendix A to make estimates of a certain efficiency parameter, our overall treatment is fairly model-independent, and can be used as a framework within which to compare various photoionization scenarios. Our basic picture is roughly the following: An ever larger fraction of the baryons in the universe falls into nonlinear structures and forms galaxies. A certain fraction of these baryons form stars or quasars which emit ultraviolet radiation, and some of this radiation escapes into the ambient intergalactic medium (IGM) and ends up photoionizing and heating it. Due to cooling losses and recombinations, the net number of ionizations per UV photon is generally less than unity.

Apart from photoionization, early galaxies can also ionize the IGM through supernova driven winds, an ionization mechanism that will not be treated in this chapter. Although such winds can ionize the IGM by $z = 5$, early enough to explain the absence of a Gunn-Peterson effect (Tegmark *et al.* 1993), the relatively low velocities of such winds makes them unable to distribute the released energy throughout space at redshifts early enough (by $z \approx 50$) to measurably affect the CBR.

Our approach will be to first write the ionization fraction of the IGM as a product of a number of factors, and then discuss the value of each of these factors in more detail. Let us write

$$\chi = f_s f_{uvpp} f_{ion}, \tag{3.1}$$

where

$$\begin{cases} \chi & = \text{fraction of IGM that is ionized,} \\ f_s & = \text{fraction of baryons in nonlinear structures,} \\ f_{uvpp} & = \text{UV photons emitted into IGM per proton in nonlinear structures,} \\ f_{ion} & = \text{net ionizations per emitted UV photon.} \end{cases}$$

Let us first consider the case where the UV photons are produced by stars, and return to the quasar case later. Using the fact that a fraction 0.0073 of the rest mass is released in stellar burning of hydrogen to helium, we obtain

$$f_{uvpp} \approx 0.0073 \left(\frac{m_p c^2}{13.6\text{eV}}\right) f_H f_{burn} f_{uv} f_{esc}, \tag{3.2}$$

where

$$\begin{cases} f_H & = \text{mass fraction hydrogen in IGM,} \\ f_{burn} & = \text{mass fraction of hydrogen burnt,} \\ f_{uv} & = \text{fraction of energy released as UV photons,} \\ f_{esc} & = \text{fraction of UV photons that escape from galaxy.} \end{cases}$$

We will take the primordial mass fraction of helium to be 24%, *i.e.* $f_H = 76\%$. Now define the *net efficiency*

$$f_{net} = f_{burn}\, f_{uv}\, f_{esc}\, f_{ion},$$

and equation (3.1) becomes

$$\chi \approx 3.8 \times 10^5 \, f_{net}\, f_s. \tag{3.3}$$

The key feature to note about this expression is that since 3.8×10^5 is such a large number, quite modest efficiencies f_{net} still allow χ to become of order unity as soon as a very small fraction of the baryons are in galaxies. As will be seen in the next section, this means that reionization is possible even at redshifts far out in the Gaussian tail of the distribution of formation redshifts, at epochs long before those when the bulk of the baryons go nonlinear. This appears to have been first pointed out by Couchman and Rees (1986).

3.2 The Mass Fraction in Galaxies

In this section, we will discuss the parameter f_s. Assuming the standard PS theory of structure formation (Press & Schechter 1974), the fraction of all mass that has formed gravitationally bound objects of total (baryonic and non-baryonic) mass greater than M at redshift z is the integral of the Gaussian tail,

$$f_s = \operatorname{erfc}\left[\frac{\delta_c}{\sqrt{2}\sigma(M,z)}\right], \tag{3.4}$$

where the complementary error function $\operatorname{erfc}(x) \equiv 2\pi^{-1/2}\int_x^\infty e^{-u^2}du$ and $\sigma(M,z)$ is the r.m.s. mass fluctuation in a sphere containing an expected mass M at redshift z. σ^2 is given by top-hat filtering of the power spectrum as

$$\sigma(M,z)^2 \propto \int_0^\infty P(k)\left[\frac{\sin kr_0}{(kr_0)^3} - \frac{\cos kr_0}{(kr_0)^2}\right]^2 dk, \tag{3.5}$$

where $P(k)$ is the power spectrum at redshift z and r_0 is given by $\frac{4}{3}\pi r_0^3\rho = M$, $\rho = \frac{3H^2\Omega}{8\pi G}$ being the density of the universe at redshift z. Although this approach has been criticized as too simplistic, numerical simulations (Efstathiou *et al.* 1988; Efstathiou & Rees 1988; Carlberg & Couchman 1989) have shown that it describes the mass distribution of newly formed structures remarkably well. Making the standard assumption of a Gaussian density field, Blanchard *et al.* (1992) have argued that it is an accurate description at least in the low mass limit. Since we are mainly interested in extremely low masses such as $10^6 M_\odot$, it appears to suffice for our purposes.

For our middle-of-the road estimate, we choose $\delta_c = 1.69$, which is the linearly extrapolated overdensity at which a spherically symmetric perturbation has collapsed into a virialized object (Gott & Rees 1975; Efstathiou *et al.* 1988; Brainerd & Villumsen 1992). We take $\delta_c = 1.44$ (Carlberg & Couchman 1989) for the optimistic estimate, although the even lower value $\delta_c = 1.33$ has been discussed (Efstathiou & Rees 1988), and $\delta_c = 2.00$ (Gelb & Bertschinger 1992) for the pessimistic estimate. (Here and throughout this chapter, parameter choices are referred to as optimistic if they permit earlier reionization.)

The fact that $\sigma(M,z) \to \infty$ as $M \to 0$ implies that if we consider arbitrarily small scales, then all dark matter is in non-linear structures. Thus if no forces other than gravity were at work, so that the baryons always followed the dark matter, we would simply have $f_s = 1$ at all z. However, it is commonly believed that galaxies correspond only to objects that are able to cool (and fragment into stars) in a dynamical time or a Hubble time (Binney 1977; Rees & Ostriker 1977; Silk 1977; White & Rees 1978). The former applies to ellipticals and bulges, the latter to disks. Let us define the *virialization redshift* $(1 + z_{vir}) \equiv (\sqrt{2}/\delta_c)\sigma(M_c, 0)$, where M_c is some characteristic cutoff mass which is the total mass (baryonic and dark) of the first galaxies to form. z_{vir} is roughly the redshift at which the bulk of all baryons goes non-linear. Using equation (3.4) and the fact that $\sigma(M,z)=\sigma(M,0)/(1+z)$ in the linear regime of CDM, we thus have

$$f_s = \operatorname{erfc}\left[\frac{1+z}{1+z_{vir}}\right]. \tag{3.6}$$

A common assumption is that $M_c \approx 10^6 M_\odot$, roughly the Jeans mass at recombination. Blanchard *et al.* (1992) examine the interplay between cooling and gravitational collapse in considerable detail, and conclude that the first galaxies to form have masses in the range $10^7 M_\odot$ to $10^8 M_\odot$, their redshift distribution still being given by equation (3.6), whereas Couchman & Rees (1986) argue that the first galaxies to form may have had masses as low as $10^5 M_\odot$.

As our CDM power spectrum today, we will use that given by BBKS (Bardeen *et al.* 1986) and an $n=1$ Harrison-Zel'dovich primordial spectrum:

$$P(k) \propto \left(\frac{q^{-1}\ln(1+2.34q)}{[1+3.89q+(16.1q)^2+(5.46q)^3+(6.71q)^4]^{1/4}}\right)^2 q,$$

where $q \equiv k/[h^2\Omega_0 \mathrm{Mpc}^{-1}]$. Throughout this chapter, we will take $\Omega_0 = 1$.

Evaluating the σ^2-integral in equation (3.5) numerically yields

$$\sigma(10^5 M_\odot, 0) \approx 33.7 b^{-1}$$

for $h = 0.8$ and

$$\sigma(10^8 M_\odot, 0) \approx 13.6 b^{-1}$$

for $h = 0.5$, where the so called bias factor $b \equiv \sigma(8h^{-1}\mathrm{Mpc}, 0)^{-1}$ has been estimated to lie between 0.8 (Smoot *et al.* 1992) and 2.5 (Bardeen *et al.* 1986). Our pessimistic, middle-of-the-road and optimistic CDM estimates of z_{vir} are given in Table 3.1, and the dependence of z_{vir} on M_c is plotted in Figure 3.1. This figure also shows three alternative models of structure formation: CDM with cosmological constant (Efstathiou *et al.* 1992); tilted CDM (Cen *et al.* 1992) and MDM, mixed hot and dark matter (Shafi & Stecker 1984; Schaefer & Shafi 1992; Davis *et al.* 1992; Klypin *et al.* 1993). For the model with cosmological constant, we have taken a flat universe with $h = 0.5$, $\Omega_0 = 0.4$ and $\lambda_0 = 0.6$. For the tilted model, the power spectrum $P(k)$ is simply multiplied by a factor k^{n-1}, where we have taken $n = 0.7$. For the tilted case, equation (3.6) still applies. For the MDM case, however, perturbations in the cold component grow slower than linearly with the scale factor $(1+z)^{-1}$ and equation (3.6) is not valid. For the low masses we are considering, we have (Bond & Szalay 1983)

$$\sigma_{MDM}(M_c, z) \approx \frac{\sigma_{MDM}(M_c,0)}{(1+z)^\alpha}, \quad \text{where}$$

$$\alpha \equiv \frac{1}{4}\left[\sqrt{25-24\Omega_{HDM}}-1\right].$$

Using the parameters from Davis *et al.* (1992), who take $\Omega = 1$ and $\Omega_{HDM} = 0.3$, the MDM version of equation (3.6) becomes

$$f_s = \operatorname{erfc}\left[\left(\frac{1+z}{1+z_{vir}}\right)^\alpha\right]$$

	Mixed	Tilted	Lambda	Pess.	Mid.	Opt.
M_c	$10^6 M_\odot$	$10^6 M_\odot$	$10^6 M_\odot$	$10^8 M_\odot$	$10^6 M_\odot$	$10^5 M_\odot$
Model	MDM	Tilted	Lambda	CDM	CDM	CDM
h	0.5	0.5	0.5	0.5	0.5	0.8
b	1	1	1	2	1	0.8
δ_c	1.69	1.69	1.69	2.00	1.69	1.44
z_{vir}	2.9	10.1	8.4	4.8	17.2	41.4

Table 3.1: Galaxy formation assumptions

	Pess.	Mid.	Opt.
f_{burn}	0.2%	1%	25%
f_{esc}	10%	20%	50%
f_{uv}	5%	25%	50%
f_{ion}	10%	40%	95%
f_{net}	1×10^{-6}	2×10^{-4}	6×10^{-2}
f_{uvpp}	4	190	24,000

Table 3.2: Efficiency parameters used

where $\alpha \approx 0.8$ and we redefine z_{vir} by

$$(1 + z_{vir}) \approx \left[\frac{\sqrt{2}\,\sigma(M_c, 0)}{5.6\delta_c}\right]^{1/\alpha},$$

with $\sigma(M_c, 0)$ referring to a pure CDM power spectrum.

For the Λ case, perturbations grow approximately linearly until the universe becomes vacuum dominated at $z \approx \Omega_0^{-1} - 1 = 1.5$, after which their growth slowly grinds to a halt. A numerical integration of the Friedmann equation and the equation for perturbation growth using $h = 0.5$, $\Omega_0 = 0.4$ and $\lambda_0 = 0.6$ gives

$$\sigma_\Lambda(M_c, z) \approx 1.2\frac{\sigma(M_c, 0)}{(1 + z)}$$

for $z \gg 3$.

Since our Λ-model yields a value of z_{vir} very similar to our tilted model, we will omit the former from future plots.

3.3 Efficiency Parameters

In this section, we will discuss the various parameters that give f_{net} when multiplied together. The conclusions are summarized in Table 3.2.

f_{burn}, the fraction of galactic hydrogen that is burnt into helium during the early life of the galaxy (within a small fraction of a Hubble after formation), is essentially the galactic metallicity after the first wave of star formation. Thus it is the product of the fraction of the hydrogen that forms stars and the average metallicity per star (weighted by mass). This depends on the stellar mass function, the galactic star formation rate and the final metallicities of the high-mass stars. For our middle-of-the-road estimate, we follow Miralda-Escudé & Ostriker (1990) in taking $f_{burn} = 1\%$, half the solar value. An upper limit to f_{burn}

UV source	Spectrum $P(\nu)$	f_{uv}	$\langle E_{uv}\rangle$	T^*	σ_{18}
O3 star	$T = 50,000$K Planck	0.57	17.3 eV	28,300K	2.9
O6 star	$T = 40,000$K Planck	0.41	16.6 eV	23,400K	3.4
O9 star	$T = 30,000$K Planck	0.21	15.9 eV	18,000K	3.9
Pop. III star	$T = 50,000$K Vacca	0.56	18.4 eV	36,900K	2.2
Black hole, QSO	$\alpha = 1$ power law		18.4 eV	37,400K	1.7
?	$\alpha = 2$ power law		17.2 eV	27,800K	2.7
?	$\alpha = 0$ power law		20.9 eV	56,300K	0.6
?	$T = 100,000$K Planck	0.89	19.9 eV	49,000K	1.6

Table 3.3: Spectral parameters

is obtained from the extreme scenario where all the baryons in the galaxy form very massive and short-lived stars with $M \approx 30M_\odot$, whose metallicity could get as high as 25% (Woosley & Weaver 1986). Although perhaps unrealistic, this is not ruled out by the apparent absence of stars with such metallicities today, since stars that massive would be expected to collapse into black holes.

In estimating f_{esc}, the fraction of the UV photons that despite gas and dust manage to escape from the galaxy where they are created, we follow Miralda-Escudé & Ostriker (1990).

For f_{uv}, the fraction of the released energy that is radiated above the Lyman limit, we also follow Miralda-Escudé & Ostriker (1990). The upper limit refers to the extreme $30M_\odot$ scenario mentioned above. For reference, the values of f_{uv} for stars with various spectra are given in Table 3.3, together with some other spectral parameters that will be defined and used in Appendix A. All these parameters involve spectral integrals, and have been computed numerically.

The parameter f_{ion} is estimated in Appendix A.

An altogether different mechanism for converting the baryons in nonlinear structures into ultraviolet photons is black hole accretion. If this mechanism is the dominant one, equation (3.2) should be replaced by

$$f_{uvpp} \approx \left(\frac{m_p c^2}{13.6\text{eV}}\right) f_{bh} f_{acc} f_{uv} f_{esc},$$

where

$$\begin{cases} f_{bh} & = \text{mass fraction of nonlinear structures that end up as black holes,} \\ f_{acc} & = \text{fraction of rest energy radiated away during accretion process,} \\ f_{uv} & = \text{fraction of energy released as UV photons,} \\ f_{esc} & = \text{fraction of UV photons that escape from host galaxy.} \end{cases}$$

There is obviously a huge uncertainty in the factor f_{bh}. However, the absence of the factor 0.0073×0.76 compared to equation (3.2) means that the conversion of matter into radiation is so much more efficient that the black hole contribution might be important even if f_{bh} is quite small. For instance, $f_{acc} = 10\%$ and $f_{esc} = 100\%$ gives $f_{uvpp} \approx 10^8 f_{bh} f_{uv}$, which could easily exceed the optimistic value $f_{uvpp} \approx 24,000$ for the stellar burning mechanism in Table 3.2.

In Figure 3.2, the ionization fraction $\chi(z)$ is plotted for various parameter values using equations (3.1) and (3.6). It is seen that the ionization grows quite abruptly, so that we may speak of a fairly well-defined *ionization redshift.* Let us define the ionization redshift z_{ion} as the redshift when χ becomes 0.5, *i.e.*

$$1 + z_{ion} = (1 + z_{vir})\mathrm{erfc}^{-1}\left(\frac{1}{2f_{uvpp}f_{ion}}\right). \tag{3.7}$$

This dependence of z_{ion} of the efficiency is shown in Figure 3.3 for our various galaxy formation scenarios. It is seen that the ionization redshift is fairly insensitive to the net efficiency, with the dependence being roughly logarithmic for $f_{net} > 0.0001$.

3.4 Scattering History

For a given ionization history $\chi(z)$, the Thomson opacity out to a redshift z, the probability that a CBR photon is Thomson scattered at least once after z, is

$$P_s(z) = 1 - e^{-\tau(z)},$$

where the optical depth for Thomson scattering is given by

$$\begin{cases} \tau(z) &= \tau^* \int_0^z \frac{1+z'}{\sqrt{1+\Omega_0 z'}}\chi(z')dz', \\ \tau^* &= \frac{3\Omega_{igm}}{8\pi}\left[1 - \left(1 - \frac{\chi_{He}}{4\chi}\right) f_{He}\right] \frac{H_0 c\sigma_t}{m_p G} \approx 0.057h\Omega_{igm}, \end{cases}$$

where we have taken the mass fraction of helium to be $f_{He} \approx 24\%$ and assumed $x_{He} \approx x$, *i.e.* that helium never becomes doubly ionized and that the fraction that is singly ionized equals the fraction of hydrogen that is ionized. The latter is a very crude approximation, but makes a difference of only 6%. We assume that $\Omega_0 = 1$ throughout this chapter. Ω_{igm} denotes the density of the intergalactic medium divided by the critical density, and is usually assumed to equal Ω_b, the corresponding density of baryons. the probability that a CBR photon is Thomson scattered at least once after the standard recombination epoch at $z \approx 10^3$.

The profile of the last scattering surface is given by the so called visibility function

$$f_z(z) \equiv \frac{dP_s}{dz}(z),$$

which is the probability distribution for the redshift at which a photon last scattered. An illuminating special case is that of complete ionization at all times, *i.e.* $\chi(z) = 1$, which yields

$$P_s(z) = 1 - \exp\left(-\frac{2}{3}\tau^*\left[(1+z)^{3/2} - 1\right]\right) \approx 1 - \exp\left[-\left(\frac{z}{92}\right)^{3/2}\right] \tag{3.8}$$

for $z \gg 1$ and $h\Omega_{igm} = 0.03$. Hence we see that in order for any significant fraction of the CBR to have been rescattered by reionization, the reionization must have occurred quite early. Figures 3.4 and 3.5 show the opacity and last-scattering profile for three different choices of $h\Omega_{igm}$. In the optimistic case $h\Omega_{igm} = 0.1$, it is seen that even as low an ionization redshift as $z_{ion} = 30$ would give a total opacity $P_s \approx 50\%$. In Figures 3.6 and 3.7, we have replaced z by the angle subtended by the horizon radius at that redshift,

$$\theta(z) = 2\arctan\left[\frac{1}{2\left(\sqrt{1+z} - 1\right)}\right],$$

which is the largest angular scale on which Thomson scattering at z would affect the microwave background radiation. In Figure 3.7, we have plotted the angular visibility function $dP_s/d(-\theta)$ instead of dP_s/dz, so that the curves are probability distributions over angle instead of redshift.

In the *sudden approximation*, the ionization history is a step function

$$\chi(z) = \theta(z_{ion} - z)$$

for some constant z_{ion}, and as was discussed in Section 3.3, this models the actual ionization history fairly well. In this approximation, the visibility functions are identical to those in Figures 3.5 and 3.7 for $z < z_{ion}$, but vanish between z_{ion} and the recombination epoch at $z \approx 10^3$. Figure 3.8, which is in a sense the most important plot in this chapter, shows the total opacity $P_s(z_{ion})$ as a function of f_{net} for a variety of parameter values, as obtained by substituting equation (3.7) into (3.8). As can be seen, the resulting opacity is relatively insensitive to the poorly known parameter f_{net}, and depends mainly on the structure formation model (*i.e.* z_{vir}) and the cosmological parameter $h\Omega_{igm}$.

Thomson scattering between CBR photons and free electrons affects not only the spatial but also the spectral properties of the CBR. It has long been known that hot ionized IGM causes spectral distortions to the CBR, known as the Sunyaev-Zel'dovich effect. A useful measure of this distortion is the Comptonization y-parameter (Kompanéets 1957; Zel'dovich & Sunyaev 1969; Stebbins & Silk 1986; Bartlett & Stebbins 1991)

$$y_c = \int \left(\frac{kT}{m_e c^2}\right) n_e \sigma_t c \, dt,$$

where the integral is to be taken from the reionization epoch to today. Let us estimate this integral by making the approximation that the IGM is cold and neutral until a redshift z_{ion}, at which it suddenly becomes ionized, and after which it is remains completely ionized with a constant temperature T. Then for $\Omega = 1$, $z_{ion} \gg 1$, we obtain

$$y_c = \left(\frac{kT}{m_e c^2}\right)\left(\frac{n_{e0}\sigma_t c}{H_0}\right)\int_0^{z_{ion}} \sqrt{1+z}dz \approx 6.4 \times 10^{-8} h\Omega_{igm} T_4 \, z_{ion}^{3/2},$$

where $T_4 \equiv T/10^4$K and n_{e0}, the electron density today, has been computed as before assuming a helium mass fraction of 24% that is singly ionized. Substituting the most recent observational constraint from the COBE FIRAS experiment, $y_c < 2.5 \times 10^{-5}$ (Mather *et al.* 1994), into this expression yields

$$z_{ion} < 554 T_4^{-2/3} \left(\frac{h\Omega_{igm}}{0.03}\right)^{-2/3},$$

so all our scenarios are consistent with this spectral constraint.

3.5 Discussion

A detailed discussion of how reionization affects the microwave background anisotropies would be beyond the scope of this chapter, so we will merely review the main features. If the microwave background photons are rescattered at a redshift z, then the fluctuations we observe today will be suppressed on angular scales smaller than the angle subtended by the horizon at that redshift. This effect is seen in numerical integrations of the linearized Boltzmann equation (*e.g.* Bond & Efstathiou 1984; Vittorio & Silk 1984), and can be simply understood in purely geometrical terms. Suppose we detect a microwave photon arriving from some direction in space. Where was it just after recombination? In the absence of reionization, it would have been precisely where it appears to be coming from, say 3000 Mpc away. If the IGM was reionized, however, the photon might have originated somewhere else, scattered off of a free electron and then started propagating towards us, so at recombination it might even have been right here. Thus to obtain the observed anisotropy, we have to convolve the anisotropies at last scattering with a window function that incorporates this smoothing effect. Typical widths for the window function appropriate to the last scattering surface range from a few arc-minutes with standard recombination to the value of a few degrees that we have derived here for early reionization models.

In addition to this suppression on sub-horizon scales, new fluctuations will be generated by the first order Doppler effect and by the Vishniac effect. The latter dominates on small angular scales and is not included in the linearized Boltzmann treatment because it is a second order effect. The current upper limit on CBR fluctuations on the 1 arcminute scale of $\Delta T/T < 9\times 10^{-6}$ (Subrahmanyan *et al.* 1993) provides an interesting constraint on reionization histories through the Vishniac effect. In fact, according to the original calculations (Vishniac 1987), this would rule out most of the reionization histories in this chapter. However, a more careful treatment (Hu *et al.* 1994) predicts a Vishniac effect a factor of five smaller on this angular scale, so all reionization histories in this chapter are still permitted.

The COBE DMR detection of $\Delta T/T$ has provided a normalization for predicting CBR anisotropies on degree scales. Several experiments are underway to measure such anisotropies, and early results that report possible detections have recently become available from experiments at the South Pole (Meinhold & Lubin 1991; Shuster *et al.* 1993) and at balloon altitudes (Devlin *et al.* 1992; Meinhold *et al.* 1993; Shuster *et al.* 1993). There is some reason to believe that these detected signals are contaminated by galactic emission. Were this the case, the inferred CBR upper limits to fluctuations on degree scales might be lower than those predicted from COBE extrapolations that adopt the scale-invariant power spectrum that is consistent with the DMR result and is generally believed to be the most appropriate choice on large scales from theoretical considerations (*e.g.* Gorski *et al.* 1993; Kashlinsky 1992). In the absence of such contaminations, the detected fluctuations in at least some degree-scale experiments are, however, consistent with the COBE extrapolation (*e.g.* Jubas & Dodelson 1993). The variation from field to field, repeated on degree scales, also may argue either for galactic contamination or else for unknown experimental systematics, or even non-Gaussian fluctuations. The results of other recent experiments such as ARGO (de Bernardis *et al.* 1993), PYTHON (Dragovan *et al.* 1993) and MSAM (Cheng *et al.* 1993) have reinforced the impression that the experimental data is not entirely self-consistent, and that some form of systematic errors may be important.

The controversy over the interpretation of the degree-scale CBR fluctuations makes our reanalysis of the last scattering surface particularly timely. We have found that canonical dark matter, tailored to provide the 10 degree CBR fluctuations detected by the COBE DMR experiment, results in sufficiently early reionization (before $z \approx 50$) over a fairly wide range of parameter space, to smooth out primordial degree-scale fluctuations. Our middle-of-the-road model produces suppression by roughly a factor of two; it is difficult, although not impossible, to obtain a much larger suppression. This smoothing, because it is of order unity in scattering optical depth, is necessarily inhomogeneous. We predict the presence of regions with large fluctuations and many "hot spots" and "cold spots", corresponding to "holes" in the last-scattering surface, as well as regions with little small-scale power where the last scattering is more efficient. The detailed structure of the CBR sky in models with reionization will be left for future studies. Here we simply conclude by emphasizing that anomalously low values of $\Delta T/T$ over degree scales are a natural corollary of reionization at high redshift.

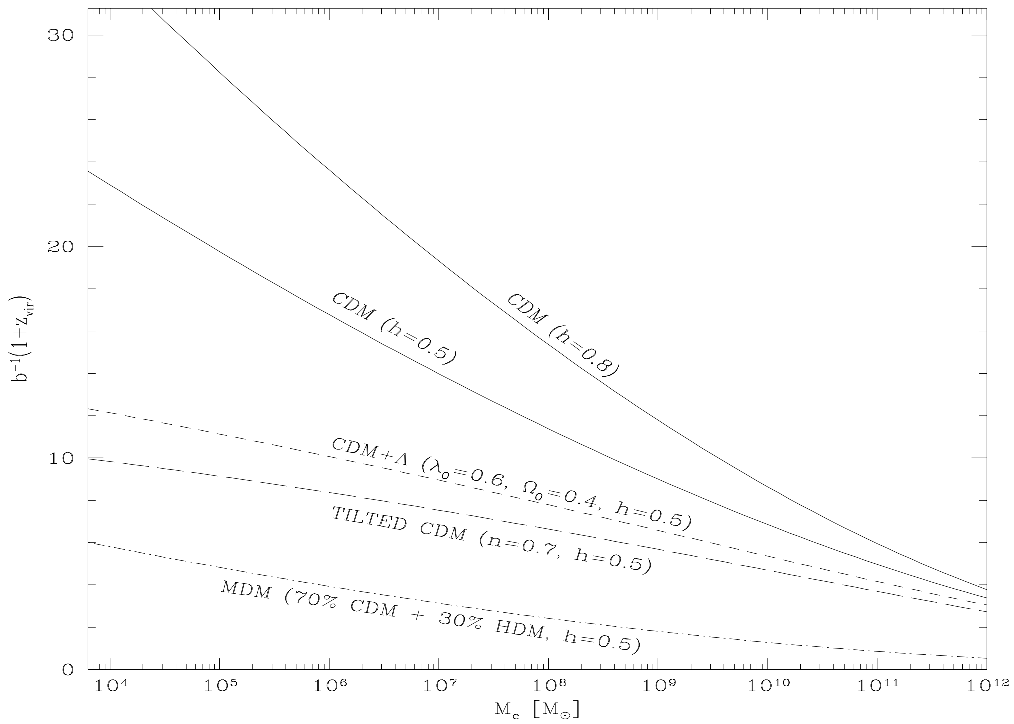


Figure 3.1: Virialization redshifts for objects of various masses.

The virialization redshift, the redshift at which the bulk of the objects of mass M_c form, is plotted for a number of cosmological models. In all cases shown, $\Omega_0 + \lambda_0 = 1$ and $\delta_c = 1.69$.

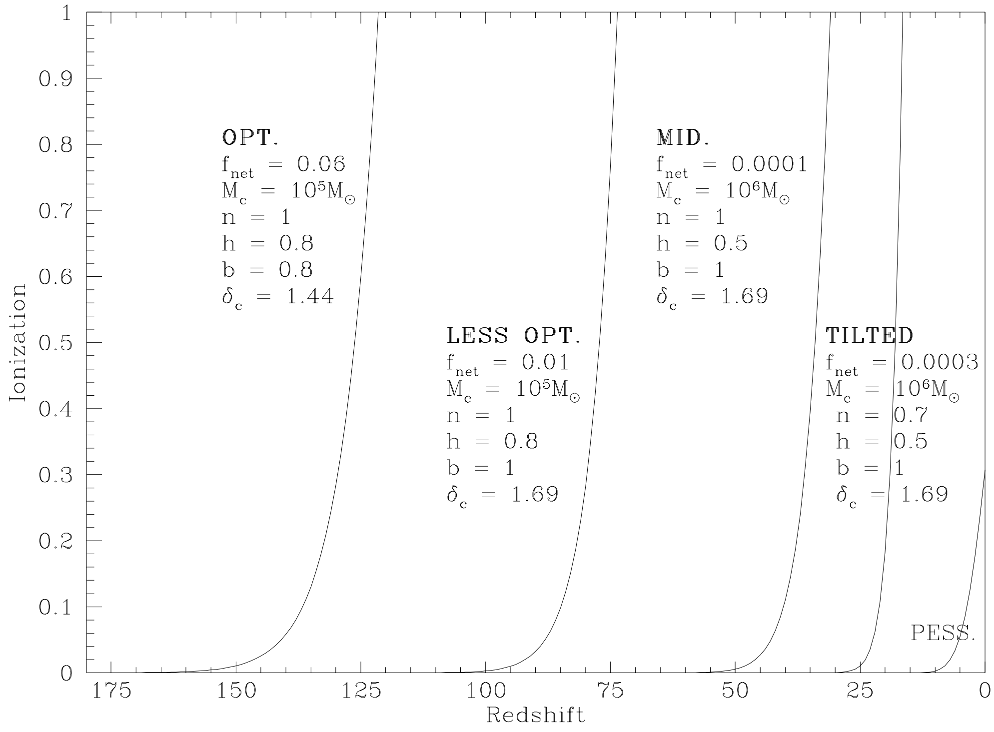


Figure 3.2: Volume fraction ionized for various scenarios.

The volume fraction of the universe that is in ionized Strömgren bubbles is plotted as a function of redshift for various parameter choices, corresponding to $n = 1$ CDM (optimistic, less optimistic, middle-of-the-road and pessimistic cases) and the tilted power spectrum (n=0.7) variant of CDM.

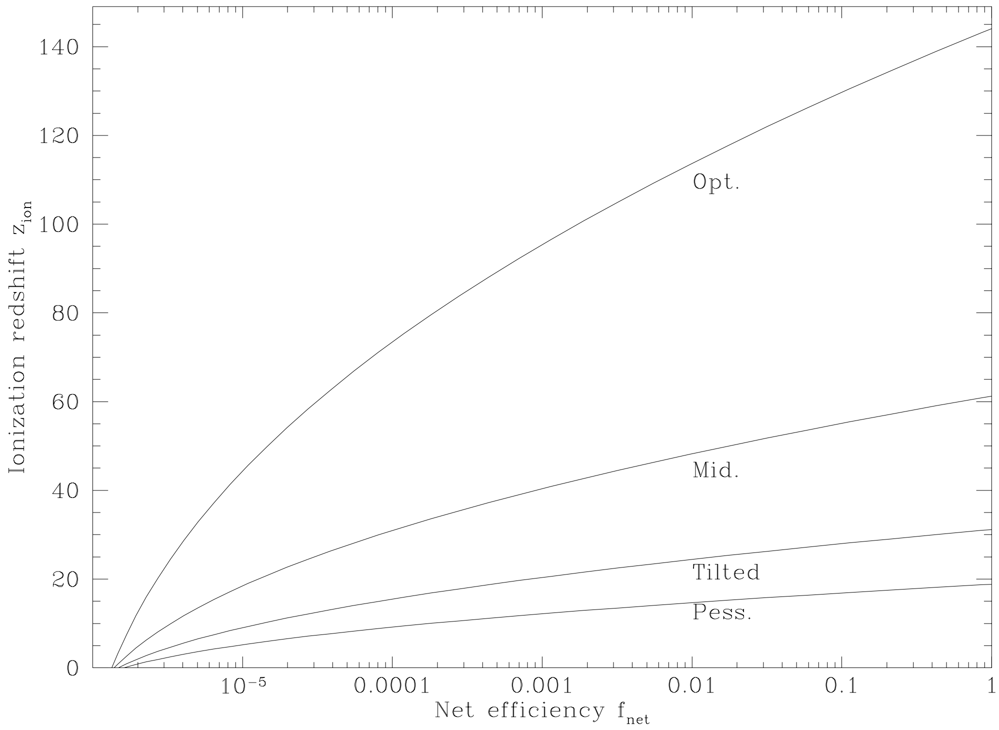


Figure 3.3: Ionization redshift for various scenarios.

The redshift at which $x = 0.5$ plotted as a function of the net efficiency. The four curves correspond to four of the choices of z_{vir} in Table 3.1: 41.4, 17.2, 8.4 and 4.8 from top to bottom.

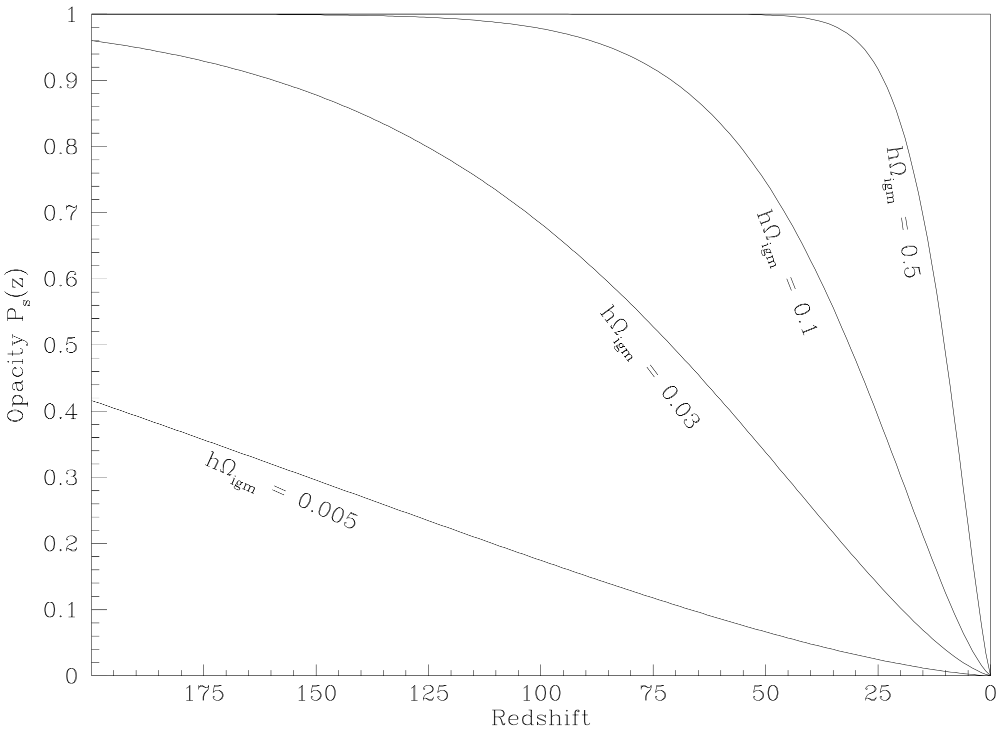


Figure 3.4: Opacity for completely ionized IGM.

The Thomson opacity $P_s(z)$, the probability that a CBR photon has been scattered at least once after the redshift z, is plotted for four different choices of $h\Omega_{igm}$ for the case where the IGM is completely ionized at all times. For more realistic scenarios where ionization occurs around some redshift z_{ion}, the opacity curves simply level out and stay constant for $z \gg z_{ion}$.

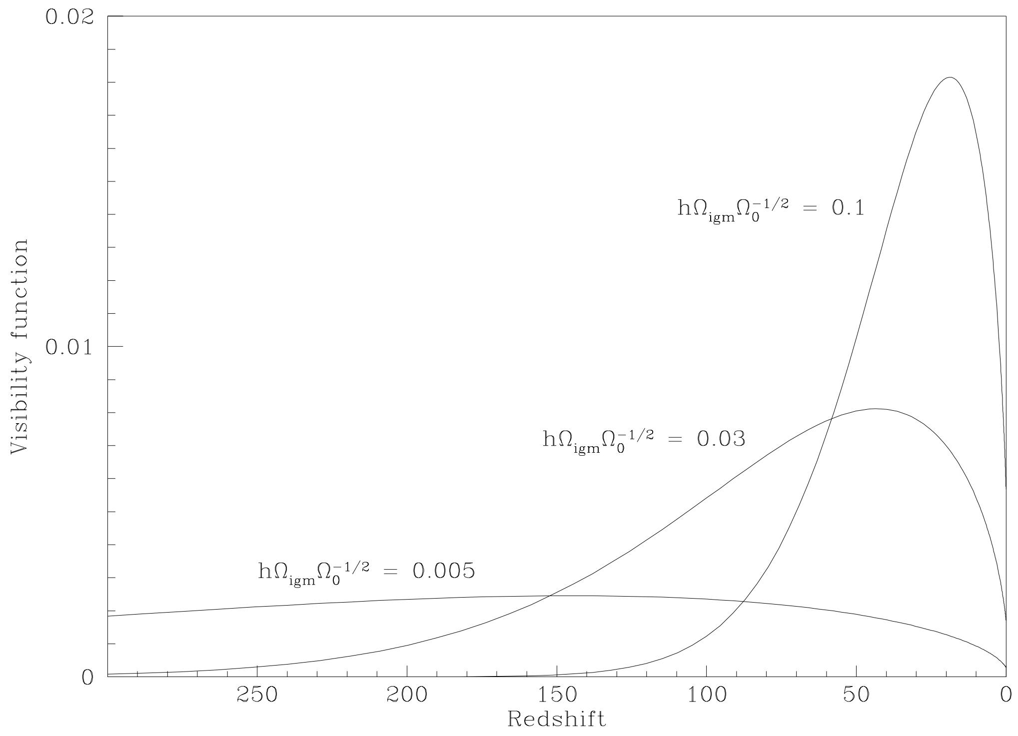


Figure 3.5: Last-scattering surface for completely ionized IGM.

The probability distribution for the redshift at which a CBR photon was last scattered, the so called visibility function, is plotted for four different choices of $h\Omega_{igm}$ for the case where the IGM is completely ionized at all times. For more realistic scenarios where ionization occurs around some redshift z_{ion}, the curves are unaffected for $z \ll z_{ion}$, vanish for $z_{ion} \ll z \ll 10^3$ and have a second bump around $z \approx 10^3$.

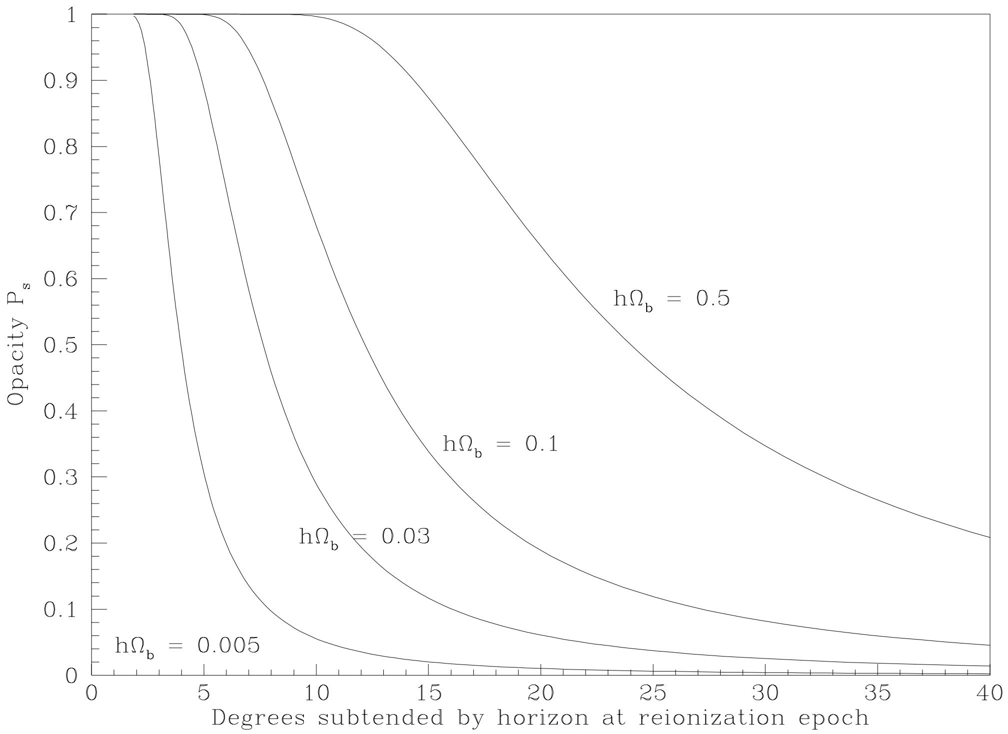


Figure 3.6: Opacity for completely ionized IGM as function of angle.

The total Thomson opacity P_s, the probability that a CBR photon has been scattered at least once since the recombination epoch, is plotted as a function of the angle in the sky that the horizon subtended at the reionization epoch. This is the largest angular scale on which fluctuations can be suppressed.

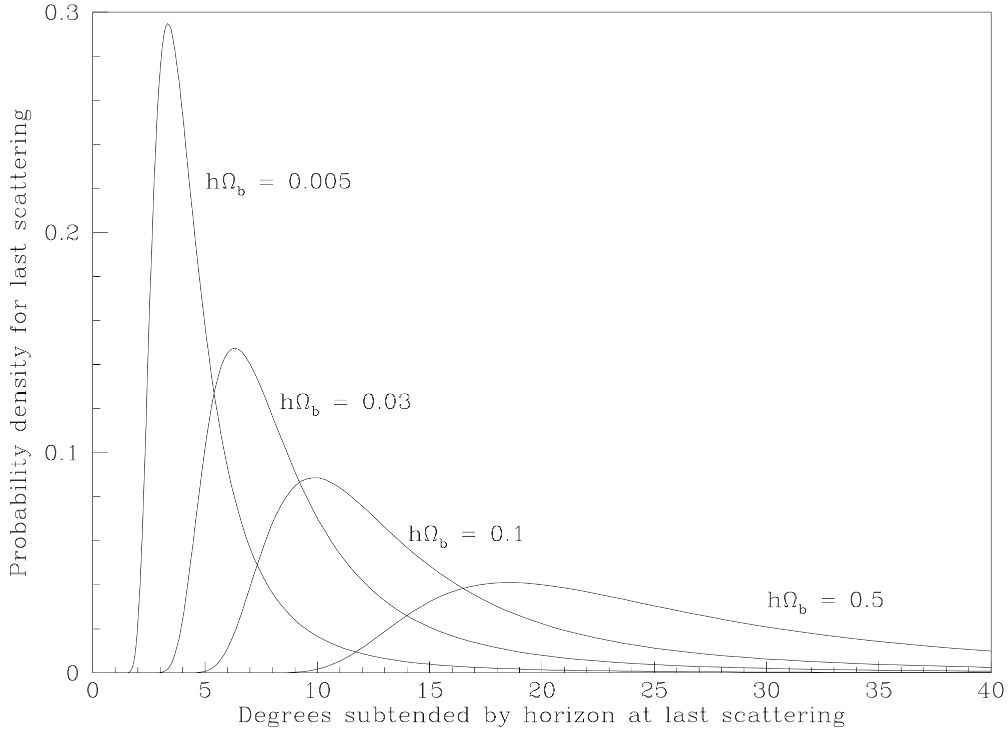


Figure 3.7: Last-scattering surface for completely ionized IGM as function of angle.

The probability distribution for the angle subtended by the horizon when a CBR photon was last scattered, the so angular visibility function, is plotted for four different choices of $h\Omega_{igm}$ for the case where the IGM is completely ionized at all times. For more realistic scenarios where ionization occurs around some redshift z_{ion}, corresponding to an angle θ_{ion}, the curves are unaffected for $\theta \gg \theta_{ion}$, vanish for $2^\circ \ll \theta \ll \theta_{ion}$ and have a second bump around $\theta \approx 2^\circ$, the horizon angle at recombination.

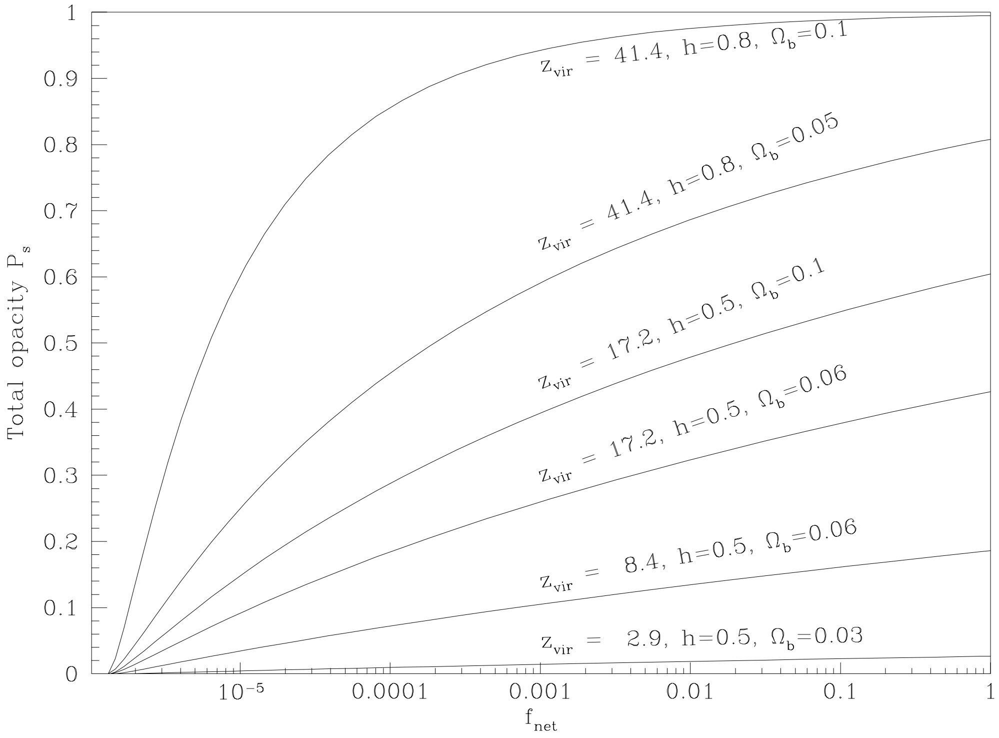


Figure 3.8: Total opacity for various models.

The total opacity, the probability that a CBR photon has been scattered at least once since the recombination epoch, is plotted for a variety of models.

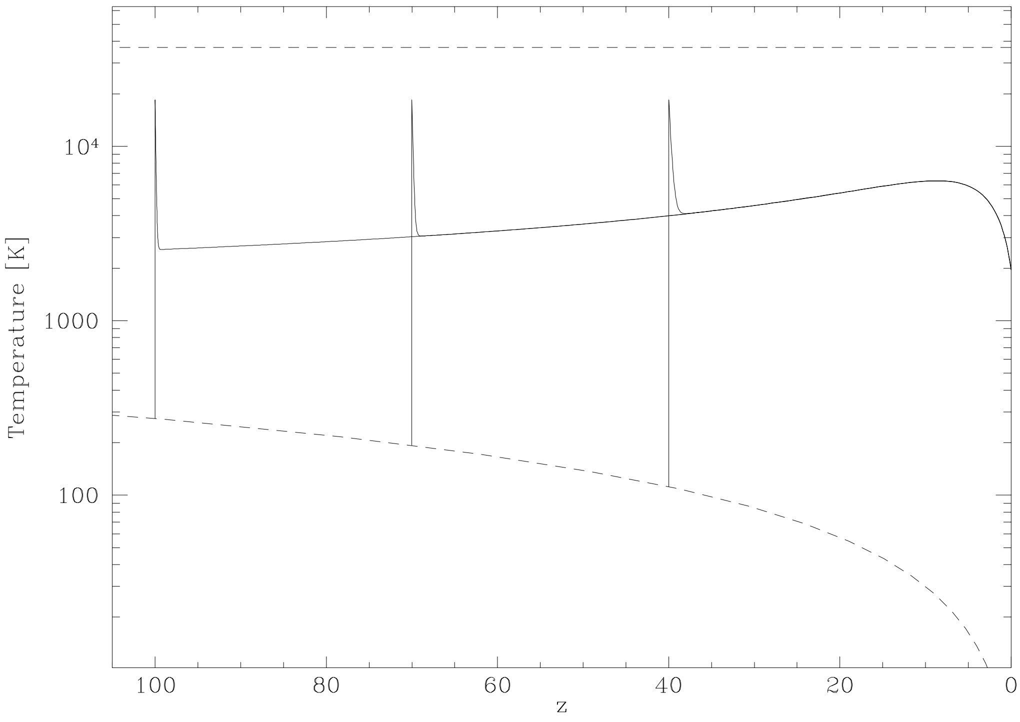


Figure 3.9: Temperature evolution in intergalactic Strömgren bubbles.

> The temperature evolution is plotted for IGM exposed to a UV flux strong enough to keep it completely photoionized. In this example, $h = 0.5$, $\Omega_{igm} = 0.06$, and $T^* = 36,900$. The upper dashed line is T^*, the temperature corresponding to the average energy of the released photoelectrons, towards which the plasma is driven by recombinations followed by new photoionizations. The lower dashed line is the temperature of the CBR photons, towards which the plasma is driven Compton cooling. The three solid curves from left to right correspond to three different redshifts for becoming part of a Strömgren bubble. The first time the hydrogen becomes ionized, its temperature rises impulsively to $T^*/2$. After this, Compton cooling rapidly pushes the temperature down to a quasi-equilibrium level, where the Compton cooling rate equals the recombination heating rate.

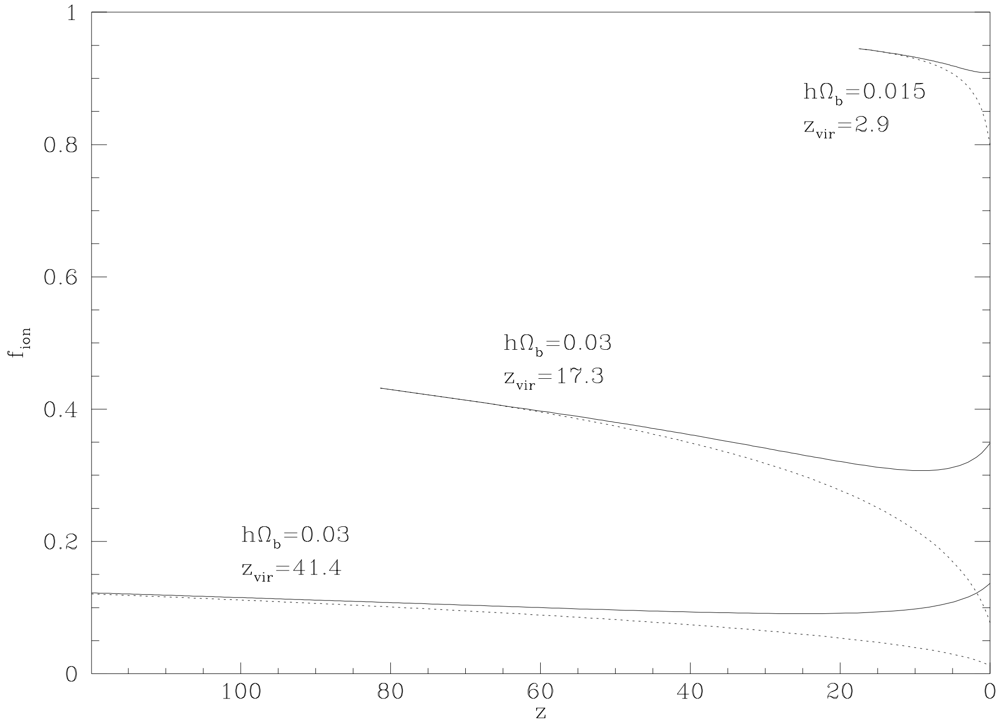


Figure 3.10: Ionization efficiencies for various scenarios

The ionization efficiency, the fraction of the UV photons that produce a net ionization, is plotted for three different parameter combinations. In all cases, $T^* = 36,900\text{K}$, the value appropriate for the radiation from the population 3 star in Table 3.3. The solid lines are the exact results from numerical integration of equation (A.5). The dotted lines are the analytic fits, which are seen to agree well in the redshift range of interest, which is typically z twice or three times z_{vir}.

Chapter 4

Reionization in an Open Universe

In this chapter, we generalize the results of the previous chapter to CDM models with $\Omega < 1$. Such models have received recent interest because the excess power in the large–scale galaxy distribution is phenomenologically fit if the "shape parameter" $\Gamma = h\Omega_0 \approx 0.25$. It has been argued that essentially all CDM models may require early reionization to suppress degree-scale anisotropies in order to be consistent with experimental data, if the lowest degree-scale measurements are indeed characteristic of the primordial temperature anisotropies. It is found that if the cosmological constant $\lambda = 0$, the extent of this suppression is quite insensitive to Ω_0, as opposing effects partially cancel. Given a σ_8-normalization today, the loss of small–scale power associated with a lower Ω_0 is partially canceled by higher optical depth from longer lookback times and by structures forming at higher redshifts as the universe becomes curvature–dominated at $z \approx \Omega_0^{-1}$. The maximum angular scale on which fluctuations are suppressed decreases when Ω_0 is lowered, but this effect is also rather weak and unlikely to be measurable in the near future. For flat models, on the other hand, where $\lambda_0 = 1 - \Omega_0$, the negative effects of lowering Ω_0 dominate, and early reionization is not likely to play a significant role if $\Omega_0 \ll 1$. The same applies to for CDM models where the Γ is lowered by increasing the number of relativistic particle species.

4.1 Introduction

The inference from some, but not all, cosmic microwave background anisotropy experiments, is that many models, including the standard cold dark matter (CDM) model, may produce excessive temperature fluctuations on degree and sub–degree angular scales (Gorski *et al.* 1993; Vittorio & Silk 1992). Early reionization at redshift $z \gtrsim 30$ produces an optical depth to scattering $\tau \gtrsim 20\%$, and suffices, with the known baryon density from primordial nucleosynthesis constraints, to reconcile CDM with all observational limits (Sugiyama *et al.* 1993). In Chapter 3, early photoionization of the intergalactic medium was discussed in a fairly model–independent way, in order to investigate whether early structures corresponding to rare Gaussian peaks in a CDM model indeed could photoionize the intergalactic medium sufficiently early to appreciably smooth out the microwave background fluctuations. In this chapter, the results of Chapter 3 will be generalized to $\Omega < 1$ models with non–zero cosmological constant Λ. Essentially all the notation used in this chapter was defined in

Chapter 3 and, in the interest of brevity, some of the definitions will not be repeated here.

Just as in Chapter 3, our basic picture is the following: An ever larger fraction f_s of the baryons in the universe falls into nonlinear structures and forms galaxies. A certain fraction of these baryons form stars or quasars which emit ultraviolet radiation. Some of this radiation escapes into the ambient intergalactic medium (IGM), which is consequently photoionized and heated. Due to cooling losses and recombinations, the net number of ionizations per UV photon, f_{ion}, is generally less than unity.

The results that we present here generalize the previous work to the case of an open universe. Lowering Ω_0 has four distinct effects:

1. Density fluctuations gradually stop growing once $z \lesssim \Omega_0^{-1}$. Thus given the observed power spectrum today, a lower Ω_0 implies that the first structures formed earlier.

2. Matter-radiation equality occurs later, which shifts the turning-point of the CDM power spectrum toward larger scales. This means less power on very small scales (such as $\sim 10^6 M_\odot$) relative to the scales at which we normalize the power spectrum (namely galaxy cluster scales, $\sim 8h^{-1}$Mpc or even the much larger COBE scale). One consequence is that the first structures form later.

3. The lookback time to a given ionization redshift becomes larger, resulting in a higher optical depth.

4. The horizon at a given ionization redshift subtends a smaller angle on the sky, thus lowering the angular scale below which CBR fluctuations are suppressed.

Thus in terms of the virialization redshift z_{vir} defined in Chapter 3, the redshift at which typical structures go nonlinear, effect 1 increases z_{vir} whereas effect 2 decreases z_{vir}. So these two effects influence f_s, the fraction of baryons in nonlinear structures, in opposite directions. As to effect 2, it should be noted that this applies not only to CDM, but to any spectrum that "turns over" somewhere between the very smallest nonlinear scales ($\sim 10^6 M_\odot$) and the very largest ($\sim 10^{21} M_\odot$) scales at which we COBE-normalize. Yet another effect of lowering Ω_0 is that the ionization efficiency f_{ion} drops slightly, at most by a factor $\Omega_0^{1/2}$. This is completely negligible compared to the above-mentioned effects, as z_{ion} depends only logarithmically on the efficiency.

In the following sections, we will discuss each of these four effects in greater detail, and then compute their combined modification of the ionization history in a few scenarios.

4.2 The Boost Factor

When curvature and vacuum density are negligible, sub-horizon-sized density fluctuations simply grow as the scale factor $a \propto (1+z)^{-1}$. Thus at early times $z \gg \Omega_0^{-1}$, we can write

$$\delta = \frac{B(\Omega_0, \lambda_0)}{1+z}\delta_0$$

for some function B independent of z that we will refer to as the *boost factor.* Clearly $B(1,0)=1$. Thus if certain structures are assumed to form when δ equals some fixed value, then given the observed power spectrum today, the boost factor tells us how much earlier these structures would form than they would in a standard flat universe. The boost factor is simply the inverse of the so called growth factor, and can be computed analytically for a number of special cases (see *e.g.* Peebles 1980). For the most general case, the fit

$$B(\Omega_0,\lambda_0)\approx\frac{2}{5\Omega_0}\left[\Omega_0^{4/7}-\lambda_0+\left(1+\frac{\Omega_0}{2}\right)\left(1+\frac{\lambda_0}{70}\right)\right]$$

is accurate to within a few percent for all parameter values of cosmological interest (Carroll *et al.* 1992). The exact results are plotted in Figure 4.1 for the case $\lambda_0=0$ and the flat case $\lambda_0=1-\Omega_0$. Since we will limit ourselves to these two cases, the simple power-law fits

$$\begin{cases} B(\Omega_0,0) & \approx\Omega_0^{-0.63},\\ B(\Omega_0,1-\Omega_0) & \approx\Omega_0^{-0.21},\end{cases}$$

which are accurate to within 1% for $0.2\leq\Omega_0\leq 1$, will suffice for our purposes. Note that the standard rule of thumb that perturbations stop growing at $1+z\approx\Omega_0^{-1}$, indicating $B\propto\Omega_0^{-1}$, is not particularly accurate in this context.

4.3 The Power Spectrum Shift

As mentioned above, lowering $h\Omega_0$ causes the first structures go nonlinear at a later redshift. This is quantified in the present section.

The standard CDM model with power-law initial fluctuations proportional to k^n predicts a power spectrum that is well fitted by (Bond & Efstathiou 1984; Efstathiou *et al.* 1992)

$$P(k)\propto\frac{q^n}{\left(1+[aq+(bq)^{1.5}+(cq)^2]^{1.13}\right)^{2/1.13}},$$

where $a\equiv 6.4$, $b\equiv 3.0$, $c\equiv 1.7$, $q\equiv(1h^{-1}\mathrm{Mpc})k/\Gamma$ and the "shape parameter" Γ will be discussed further on. Although this fit breaks down for scales comparable to the curvature scale $r_{curv}=k_{curv}^{-1}=H_0^{-1}|1-\Omega_0|^{-1/2}$, it is quite accurate for the much smaller scales that will be considered in the present chapter. Rather, its main limitation is that it breaks down if Ω_0 is so low that the baryon density becomes comparable to the density of cold dark matter. Thus for $\Omega_{igm}\approx 0.05$, the results cannot be taken too seriously for $\Omega_0<0.2$. We will limit ourselves to the standard $n=1$ model here, as the tilted ($n<1$) case was treated in Chapter 3 and was seen to be be essentially unable to reionize the universe early enough to be relevant to CBR anisotropies. The same applies to models with mixed hot and cold dark matter.

Let us define the *amplitude ratio*

$$R(\Gamma,r_1,r_2)\equiv\frac{\sigma(r_1)}{\sigma(r_2)},$$

where $\sigma(r_1)$ and $\sigma(r_2)$ are the r.m.s. mass fluctuation amplitudes in spheres of radii r_1 and r_2, *i.e.*

$$\sigma(r)^2\propto\int_0^\infty P(k)\left[\frac{\sin kr}{(kr)^3}-\frac{\cos kr}{(kr)^2}\right]^2 dk.$$

As in Chapter 3, we normalize the power spectrum so that $\sigma(8h^{-1}\mathrm{Mpc})$ equals some constant denoted σ_8, and M_c will denote the characteristic mass of the first galaxies to form. The corresponding comoving length scale r_c is given by $M_c = \frac{4}{3}\pi r_c^3\rho$, where ρ is the mean density of the universe. Thus given σ_8, what is relevant for determining when the first galaxies form is the amplitude ratio

$$R(\Gamma, r_c, 8h^{-1}\mathrm{Mpc}).$$

This ratio is computed numerically, and the results are plotted as a function of Γ in Figure 4.2 for a few different values of the cutoff mass M_c. It is easy to see why the amplitude ratio increases with Γ, since on a logarithmic scale, a decrease in Γ simply shifts the entire power spectrum towards lower k, thus decreasing the amount of power on very small scales relative to that on large scales. The fit

$$R(\Gamma, r_c, 8h^{-1}\mathrm{Mpc}) \approx 3 + 7.1\ln(1h^{-1}\mathrm{Mpc}/r_c)\Gamma$$

is accurate to within 10% for $0.05 < \Gamma < 2$ and $100\mathrm{pc} < r_c < 100\mathrm{kpc}$.

4.4 The Optical Depth

Since a lower Ω_0 implies a larger $|dt/dz|$ and an older universe, the optical depth out to a given ionization redshift z_{ion} is greater for small Ω_0. For a given ionization history $\chi(z)$, the optical depth for Thomson scattering is given by

$$\begin{cases} \tau(z) &= \tau^* \int_0^z \frac{(1+z')^2}{\sqrt{\lambda_0+(1+z')^2(1-\lambda_0+\Omega_0 z')}}\chi(z')dz', \\ \tau^* &= \frac{3\Omega_{igm}}{8\pi}\left[1-\left(1-\frac{\chi_{He}}{4\chi}\right)f_{He}\right]\frac{H_0 c\sigma_t}{m_p G} \approx 0.057h\Omega_{igm}, \end{cases}$$

where we have taken the mass fraction of helium to be $f_{He} \approx 24\%$ and assumed $\chi_{He} \approx \chi$, *i.e.* that helium never becomes doubly ionized and that the fraction that is singly ionized equals the fraction of hydrogen that is ionized. The latter is a very crude approximation, but has the advantage that the error can never exceed 6%. If the universe is fully ionized for all redshifts below z, the integral can be done analytically for $\lambda_0 = 0$:

$$\tau(z) = \frac{2\tau^*}{3\Omega_0^2}\left[2 - 3\Omega_0 + (\Omega_0 z + 3\Omega_0 - 2)\sqrt{1+\Omega_0 z}\right] \approx 0.038\frac{h\Omega_{igm}z^{3/2}}{\Omega_0^{1/2}} \qquad (4.1)$$

for $z \gg \Omega_0^{-1}$. As is evident from the asymptotic behavior of the integrand, τ is independent of λ_0 in the high redshift limit. Thus optical depth of unity is attained if reionization occurs at

$$z \approx 92\left(\frac{h\Omega_{igm}}{0.03}\right)^{-2/3}\Omega_0^{1/3}.$$

4.5 The Angular Scale

It is well known that reionization suppresses CBR fluctuations only on angular scales below the horizon scale at last scattering. Combining the standard expressions for horizon radius (*e.g.* Kolb & Turner 1990) and angular size, this angle is given by

$$\theta = 2\tan^{-1}\left[\frac{\sqrt{1+z}\Omega_0^{3/2}/2}{\Omega_0 z - (2-\Omega_0)(\sqrt{1+\Omega_0 z}-1)}\right]. \tag{4.2}$$

For $z \gg \Omega_0^{-1}$, this reduces to

$$\theta \approx \sqrt{\frac{\Omega_0}{z}}, \tag{4.3}$$

but as is evident from Figure 4.3, this is quite a bad approximation except for $z \gg 100$. If we substitute it into equation (4.1) nonetheless, to get a rough estimate, we conclude that optical depth unity is obtained at an epoch whose horizon scale subtends the angle

$$\theta \approx 12^\circ \left(\frac{h\Omega_{igm}\Omega_0}{0.03}\right)^{1/3}, \tag{4.4}$$

i.e., the dependence on all three of these cosmological parameters is relatively weak. As discussed in Chapter 3, fluctuations on angular scales much smaller than this are suppressed by a factor

$$P(z) \equiv 1 - e^{-\tau(z)},$$

the *opacity*, which is the probability that a photon was Thompson scattered after redshift z. Its derivative, the *visibility function* $f_z = dP/dz$, is the probability distribution for the redshift at which last scattering occurred, the profile of the last scattering surface. The *angular visibility function*

$$f_\theta(\theta) = \frac{dP_s}{d\theta} = \left(\frac{d\theta}{dz}\right)^{-1}\frac{dP_s}{dz}$$

is plotted in Figure 4.4 for the case where the universe never recombines (the curves for the more general case with reionization at some redshift z_{ion} can be read off from Figure 4.4 as described in the previous chapter). These functions give a good idea of the range of angular scales on which suppression starts to become important. In plotting these curves, the exact expression (4.2) has been used, rather than the approximation (4.3). It is seen that the qualitative behavior indicated by equation (4.4) is correct: as Ω_0 is lowered, the peak shifts down toward smaller angular scales, but the Ω_0-dependence is quite weak.

4.6 Cosmological Consequences

We will now compare the effect of lowering Γ in three cosmological models. The first model, which will be referred to as "open CDM" for short, has $\lambda_0 = 0$. The second model, referred to as "ΛCDM", has $\lambda_0 = 1 - \Omega_0$. The *shape parameter* essentially tells us how early the epoch of matter-radiation equality occurred, and is given by

$$\Gamma = h\Omega_0\left(\frac{g_*}{3.36}\right)^{-1/2},$$

where $g_* = 3.36$ corresponds to the standard model with no other relativistic degrees of freedom than photons and three massless neutrino species. In open CDM and ΛCDM, we have $\gamma_* = 3.36$, so that $\Omega_0 = \Gamma/h$. The third model, referred to as τCDM (Dodelson, Gyuk

	Pess.	Mid.	Opt.	Very opt.
σ_8	0.5	1	1.1	1.2
δ_c	2.00	1.69	1.44	1.33
h	0.5	0.5	0.8	0.8
$M_c[M_\odot]$	10^8	10^6	10^5	10^5
$f = f_{ion} f_{uvpp}$	1	120	23,000	10^6
$\mathrm{erfc}^{-1}[1/2f]$	0.48	2.03	3.00	3.55
$h^2\Omega_{igm}$	0.010	0.013	0.015	0.020

Table 4.1: Parameters used

& Turner 1994b), has $\lambda_0 = 0$ and $\Omega = 1$, and achieves a lower value of Γ by increasing g_* instead.

Including the effect of the boost factor, equation (3.7) in Chapter 3 becomes

$$1 + z_{ion} = \frac{\sqrt{2}\sigma_8}{\delta_c} R(h\Omega_0, r_c, 8h^{-1}\mathrm{Mpc})\, B(\Omega_0, \lambda_0)\, \mathrm{erfc}^{-1}\left[\frac{1}{2 f_{uvpp} f_{ion}}\right].$$

The ionization redshift z_{ion} is plotted as a function of the shape parameter in Figure 4.5 for the various scenarios specified in Table 4.1. It is seen that for the open model, the dependence on $\Omega_0 = \Gamma/h$ is typically much weaker than the dependence on other parameters. One reason for this is that changes in the boost factor and the amplitude ratio partially cancel each other. For ΛCDM, the Ω_0-dependence is stronger, since the boost factor is weaker. In the τCDM model, the dependence on Γ/h is even stronger, as there is no boost factor whatsoever to offset the change in the amplitude ratio.

The scenarios in Table 4.1 are similar to those in Chapter 3. In the one labeled "very optimistic", the high value for f_{uvpp}, the net number of produced UV photons per proton, is obtained by assuming that the main source of ionizing radiation is black hole accretion rather than conventional stars. Note that this speculative assumption still only increases z_{ion} by $3.55/3.00 - 1 \approx 18\%$, the efficiency dependence being merely logarithmic.

Figure 4.6, in a sense the most important plot in this chapter, shows the opacity as a function of Γ/h for the various scenarios. Because of the increase in optical depth due to larger lookback times, the open model now gives slightly larger opacities for lower $\Omega_0 = \Gamma/h$. However, this dependence is seen to be quite week. For ΛCDM, where the boost factor contributes less, the net result is seen to be the opposite; a slight decrease in the opacity for lower $\Omega_0 = \Gamma/h$. For the τCDM model, where there is neither a boost factor nor an increase in the lookback time, this drop in opacity is seen to be much sharper. Note that the dependence on other uncertain parameters, summarized by the four scenarios in Table 4.1, is quite strong. Indeed, this dependence is stronger than the effect of moderate changes in Γ/h, so in the near future, it appears unlikely that opacity limits will be able to constrain the shape parameter except perhaps in the τCDM model.

The τCDM situation is summarized in Figure 4.7. To attain at least 50% opacity, $h\Omega_{igm}$ must lie above the heavy curve corresponding to the scenario in question. On the other hand, nucleosynthesis (Malaney & Mathews 1993) places a strict upper bound on this quantity if we assume that $h \geq 0.5$[1] It is seen that a shape parameter as low as $\Gamma \simeq 0.25$, which would match large-scale structure observations (Peacock & Dodds 1994), is quite difficult to reconcile with these two constraints.

4.7 Discussion

Lowering Γ is an attractive resolution of the problem that arises in reconciling the observed structure in the universe on large scales with observations on megaparsec scales. The empirical power spectrum is well fit by $\Gamma \approx 0.25$ (Peacock & Dodds 1994). Kamionkowski & Spergel (1994) have found that primordial adiabatic fluctuations in an open universe with $\Omega \approx 0.3$ are reconcilable with large-scale CBR anisotropy. On degree scales Kamionkowski, Spergel & Sugiyama (1994) require reionization with optical depth $\tau \sim 1$ in order to reconcile the low density open model with recent experimental limits, if the lowest limits are adopted. In a low Ω_0 ΛCDM model, the situation is not so critical, but reionization is required if the lowest limits (SP91) are adopted on degree scales (Gaier *et al.* 1992); $\tau \sim 0.5$ suffices however. A similar but slightly more favorable situation occurs in a τCDM model, where $\Gamma = h\Omega_0(g_*/3.36)^{-1/2}$ is reduced by increasing g_* by a factor of ~ 4, but some reionization is still required to match SP91.

We have found that reionization giving τ in the range 0.5 to 1 is readily produced and even natural in open models. This is because of the early formation of structure in combination with the increased age of the universe, effects which compensate for the flattening of the power spectrum due to the delay in matter domination. However, the ΛCDM and τCDM models with low Ω_0 fare less well in this regard, since the loss of small-scale power is not balanced by significantly earlier structure formation. With Ω_b in the range given by standard nucleosynthesis, a significant optical depth $\tau \gtrsim 0.5$ is difficult to attain in either the ΛCDM or τCDM scenarios.

[1] The τCDM model also alters the nucleosynthesis process (Dodelson, Gyuk & Turner 1994a), but this can only marginally relax the bounds unless the τ neutrino is in a mass range incompatible with τCDM (Gyuk & Turner 1994).

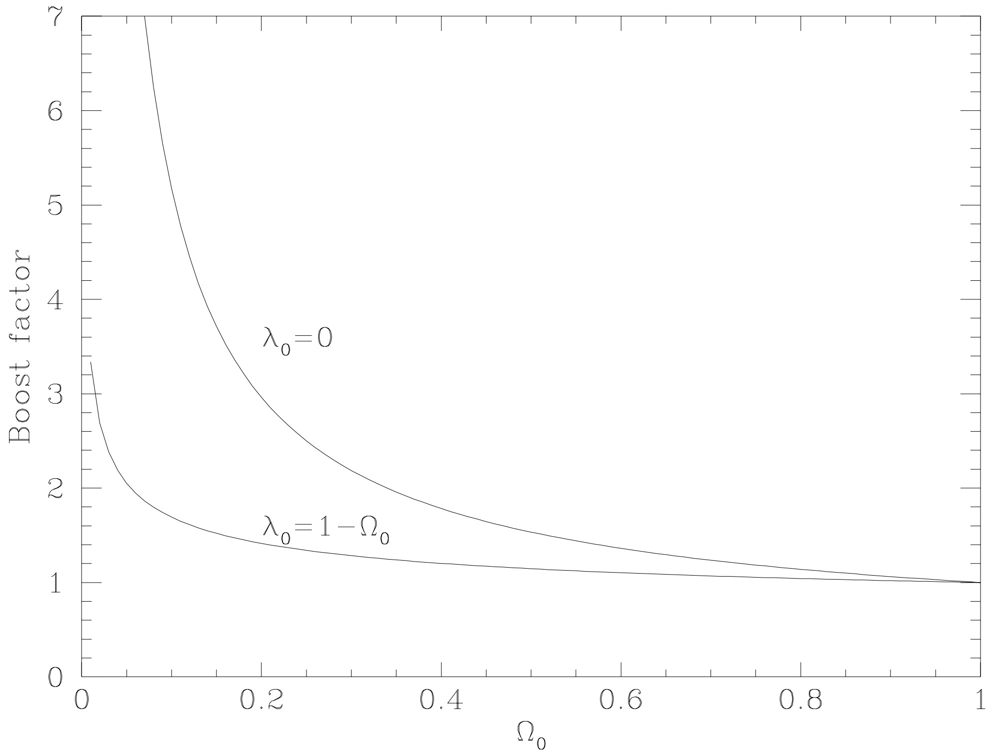


Figure 4.1: The boost factor.

The boost factor $B(\Omega_0, \lambda_0)$ is plotted as a function of Ω_0 for two classes of cosmologies. The upper curve corresponds to a standard open universe, *i.e.* $\lambda_0 = 0$, whereas the lower curve corresponds to flat universes with $\lambda_0 = 1 - \Omega_0$.

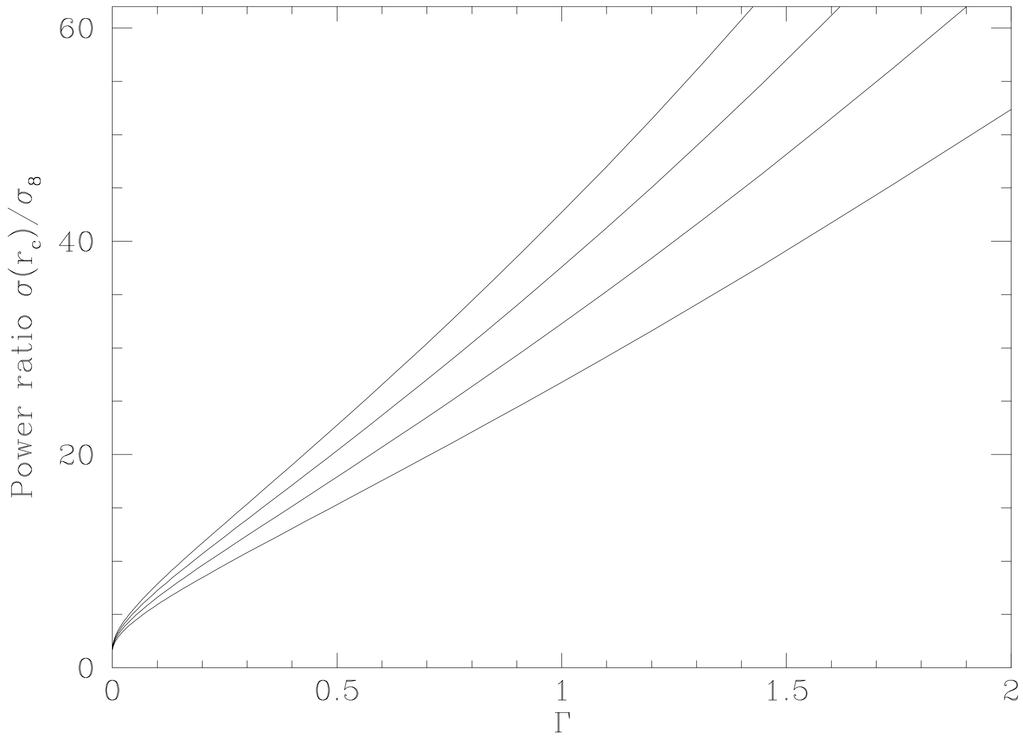


Figure 4.2: The amplitude ratio.

The ratio of the fluctuation amplitude on the small scale r_c to that at $8h^{-1}$Mpc is plotted as a function of the shape parameter Γ. From top to bottom, the four curves correspond to scales of $3.5h^{-1}$kpc, $7.5h^{-1}$kpc, $16h^{-1}$kpc and $35h^{-1}$kpc, respectively. For $h = 0.5$ and $\Omega_0 = 1$, these four length scales correspond to the masses $10^5 M_\odot$, $10^6 M_\odot$, $10^7 M_\odot$ and $10^8 M_\odot$. The weak additional dependence on Ω_0/h that would result from holding M_c rather than r_c fixed is clearly negligible, since as can be seen, M_c must vary by an entire order of magnitude to offset a mere 20% change in Γ.

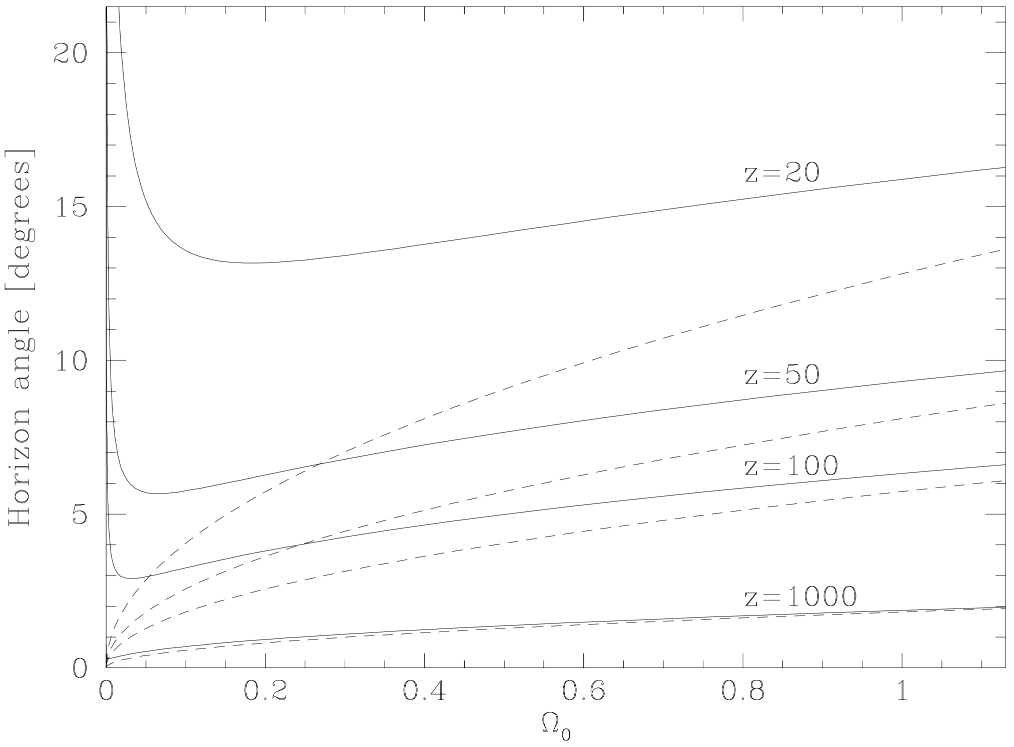


Figure 4.3: The horizon angle.

The angle in the sky subtended by a horizon volume at redshift z is plotted as a function of Ω_0 for the case with no cosmological constant. The solid lines are the exact results for the four redshifts indicated, and the dashed lines are the corresponding fits using the simplistic approximation $\theta \approx (\Omega_0/z)^{1/2}$.

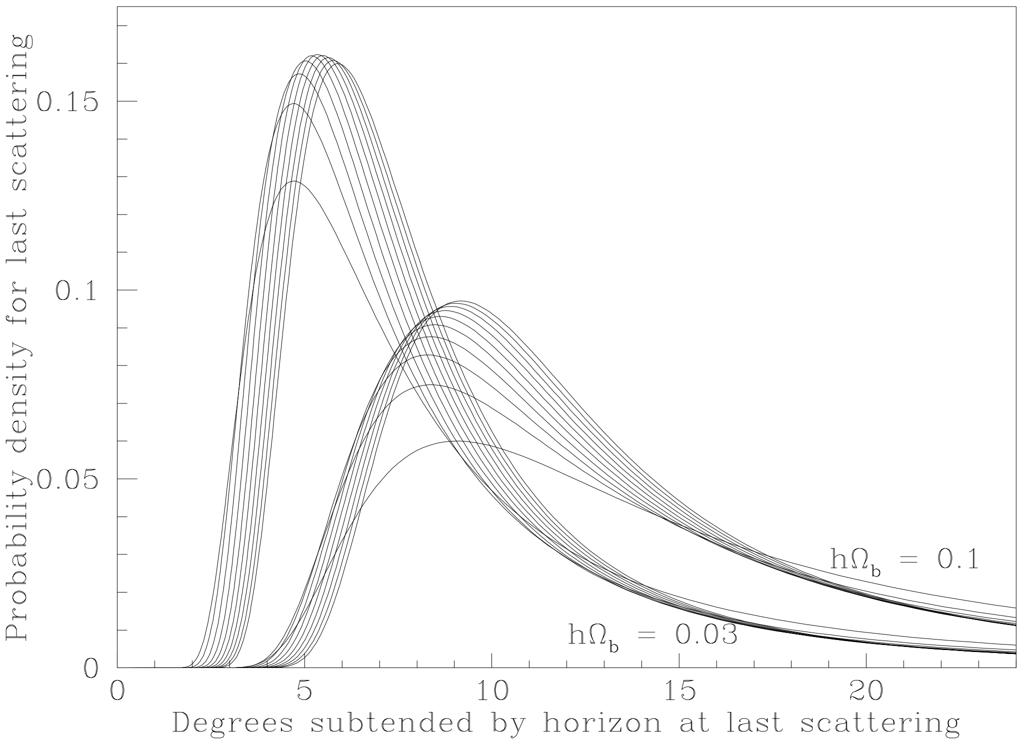


Figure 4.4: Visibility functions.

The angular visibility function for a fully ionized universe is plotted for different values of Ω_0 and diffuse baryon content $h\Omega_{igm}$. The left group of curves corresponds to $h\Omega_{igm} = 0.03$ and the right to $h\Omega_{igm} = 0.1$. Within each group, from left to right starting at the lowest peak, Ω_0 =0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 and 1.0.

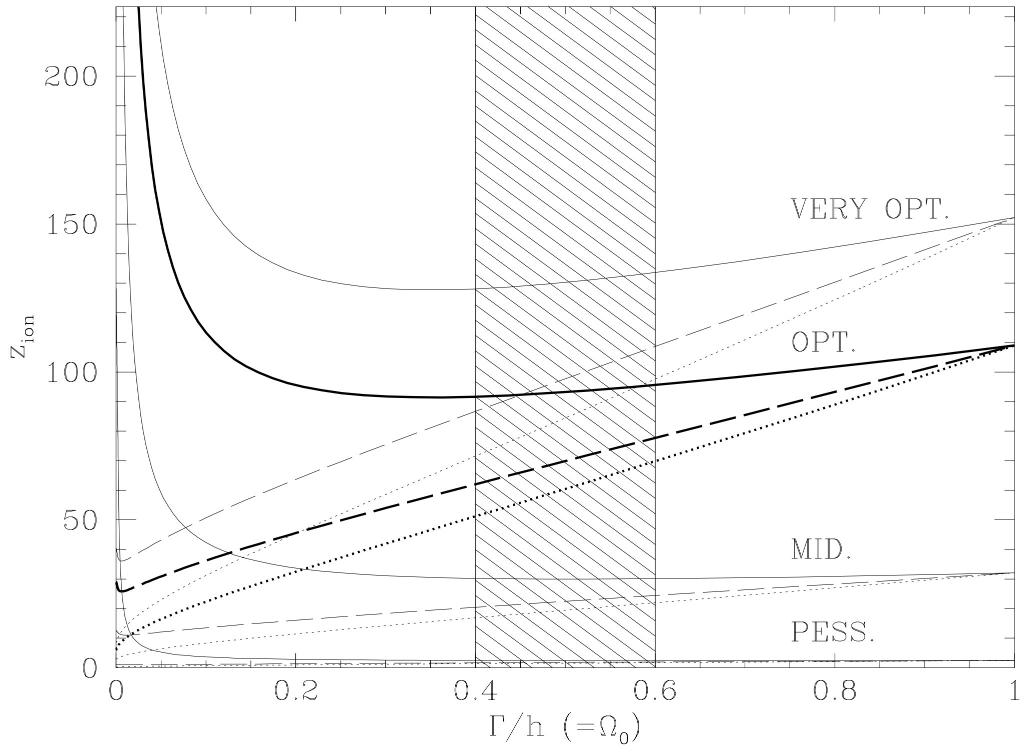


Figure 4.5: The ionization redshift.

The ionization redshift is plotted for the four scenarios described in Table 4.1. The solid lines correspond to the open case where $\lambda_0 = 0$. The dashed lines correspond to the flat case where $\lambda_0 = 1 - \Omega_0$. The dotted lines correspond to the τCDM model, where $\lambda_0 = 0$ and $\Omega_0 = 1$. Note that the combination Γ/h is really an h-independent quantity: for the open and flat cases, it is simply equal to Ω_0, and for the τCDM case it depends only on g_*, the number of relativistic particle species.

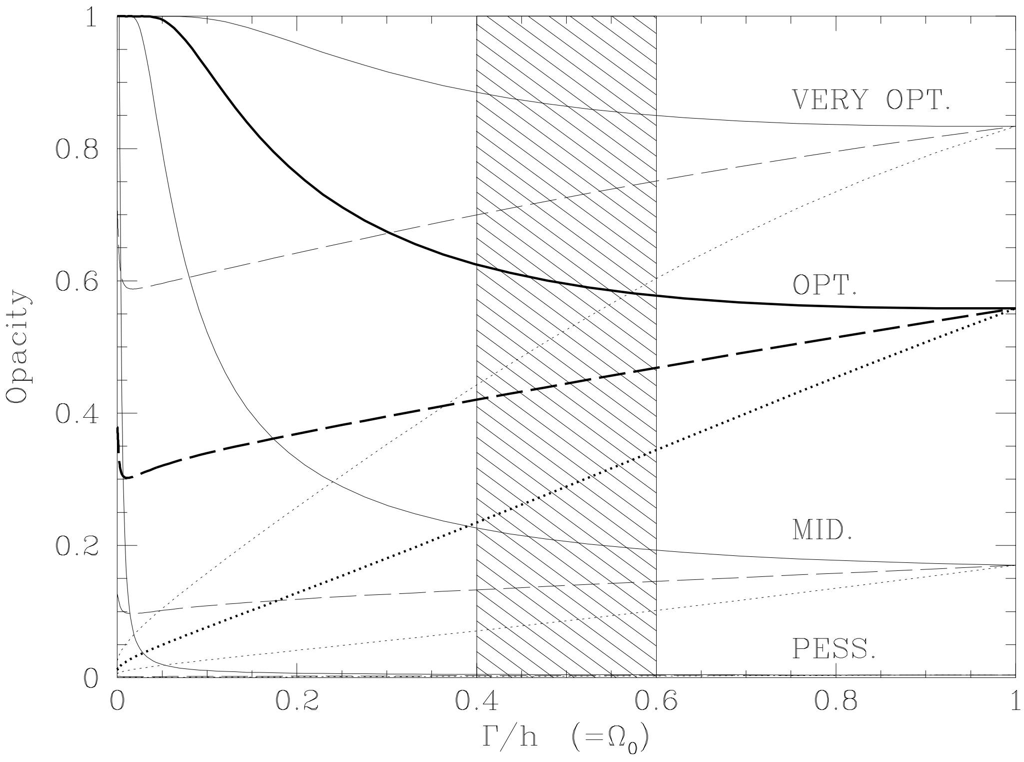


Figure 4.6: The opacity.

The opacity, the probability that a CBR photon is Thomson scattered at least once since since the standard recombination epoch, is plotted for the four scenarios described in Table 4.1. The solid lines correspond to the open case where $\lambda_0 = 0$. The dashed lines correspond to the flat case where $\lambda_0 = 1 - \Omega_0$. The dotted lines correspond to the τCDM model, where $\lambda_0 = 0$ and $\Omega_0 = 1$. Just as in Figure 4.5, note that the combination Γ/h is really an h-independent quantity: for the open and flat cases, it is simply equal to Ω_0, and for the τCDM case it depends only on g_*, the number of relativistic particle species.

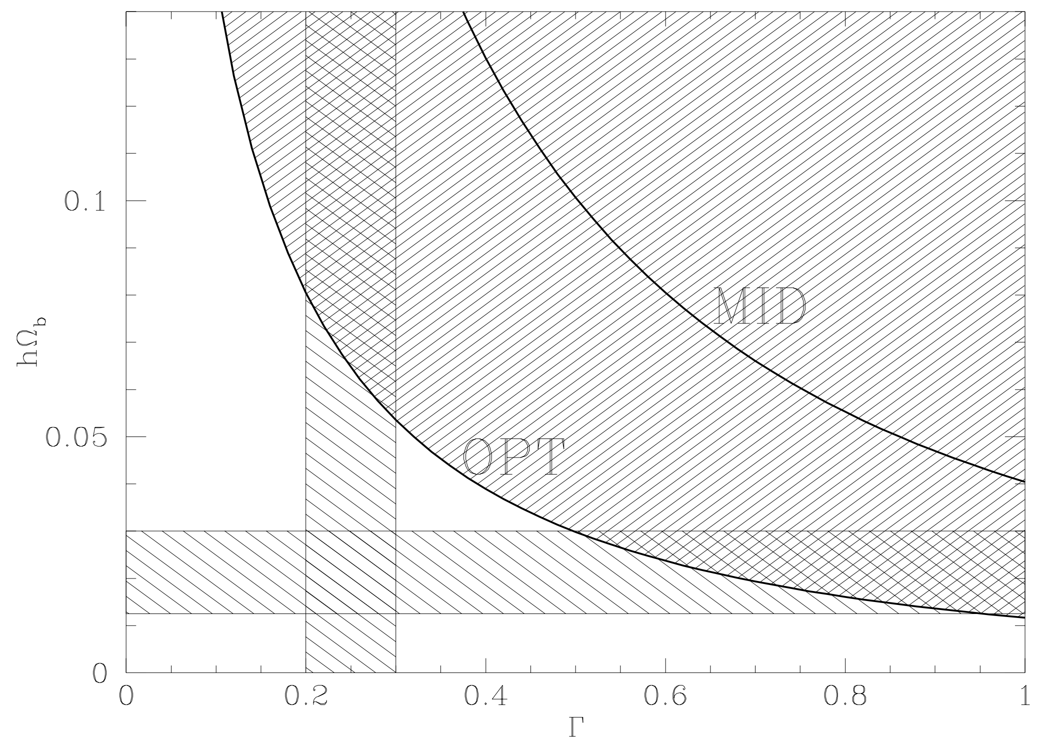


Figure 4.7: Reionization in τCDM.

The two curves show the baryon density required for 50% opacity in τCDM, *i.e.* for reionization to rescatter 50% of the CBR photons. The upper and lower heavy curves correspond to the middle-of-the road and optimistic scenarios, respectively. Thus even in the optimistic scenario, 50% opacity cannot be obtained outside of the fine-hatched region. The horizontal shaded region corresponds to the values of $h\Omega_b$ allowed by standard nucleosynthesis ($0.01 < h^2\Omega_b < 0.015$) in conjunction with the constraint $0.5 < h < 0.8$. The vertical shaded region corresponds to values of the "shape parameter" Γ preferred by power spectrum measurements. Note that these three regions do not intersect.

Chapter 5

Did the Universe Recombine?

In this chapter we will show that one still cannot conclusively assert that the universe underwent a neutral phase, despite the new COBE FIRAS limit $y < 2.5 \times 10^{-5}$ on Compton y-distortions of the cosmic microwave background. Although scenarios where the very early ($z \sim 1000$) ionization is thermal (caused by IGM temperatures exceeding 10^4K) are clearly ruled out, there is a significant loophole for cosmologies with typical CDM parameters if the dominant ionization mechanism is photoionization. If the ionizing radiation has a typical quasar spectrum, then the y-constraint implies roughly $h^{4/3}\Omega_{igm}\Omega_0^{-0.28} < 0.06$ for fully ionized models. This means that BDM models with $\Omega_0 \approx 0.15$ and reionization at $z \approx 1000$ are strongly constrained even in this very conservative case, and can survive the y test only if most of the baryons form BDM around the reionization epoch.

5.1 Introduction

Recombination of the primeval plasma is commonly assumed but was by no means inevitable. Theories exist that predict early reionization are as diverse as those invoking primordial seed fluctuations that underwent early collapse and generated sources of ionizing radiation, and models involving decaying or annihilating particles. The former class includes cosmic strings and textures, as well as primordial isocurvature baryon fluctuations. The latter category includes baryon symmetric cosmologies as well as decaying exotic particles or neutrinos.

The Compton y-distortion of the cosmic microwave background (CBR) provides a unique constraint on the epoch of reionization. In view of the extremely sensitive recent FIRAS limit of $y < 2.5 \times 10^{-5}$, we have reinvestigated constraints on the early ionization history of the intergalactic medium (IGM), and have chosen to focus on what we regard as the most important of the non-standard recombination history models, namely the primordial isocurvature baryon scenario involving a universe dominated by baryonic dark matter (BDM), as advocated by Peebles (1987); Gnedin & Ostriker (1992) (hereafter "GO"); Cen, Ostriker & Peebles (1993) and others. This class of models takes the simplest matter content for the universe, namely baryons, to constitute dark matter in an amount that is directly observed and is even within the bounds of primordial nucleosynthesis, if interpreted liberally, and can reconstruct essentially all of the observed phenomena that constrain large-scale structure. The BDM model is a non-starter unless the IGM underwent very early reionization, in order to avoid producing excessive CBR fluctuations on degree scales. Fortunately, early nonlinearity is inevitable with BDM initial conditions, $\delta\rho/\rho \propto M^{-5/12}$, corresponding to a power-spectrum $\langle\delta_k^2\rangle \propto k^{-1/2}$ for the observationally preferred choice of spectral index (Cen, Ostriker & Peebles 1993).

Is it possible that the IGM has been highly ionized since close to the standard recombination epoch at $z \approx 1100$? Perhaps the most carefully studied BDM scenario in which this happens is that by GO. In their scenario, $\Omega_0 = \Omega_{b0} \approx 0.15$. Shortly after recombination, a large fraction of the mass condenses into faint stars or massive black holes, releasing energy that reionizes the universe and heats it to $T > 10,000$K by $z = 800$, so Compton scattering off of hot electrons causes strong spectral distortions in the cosmic microwave background. The models in GO give a Compton y-parameter between 0.96×10^{-4} and 3.1×10^{-4}, and are thus all ruled out by the most recent observational constraint from the COBE FIRAS experiment, $y < 2.5 \times 10^{-5}$ (Mather *et al.* 1994).

There are essentially four mechanisms that can heat the IGM sufficiently to produce Compton y-distortions:

- Photoionization heating from UV photons (Shapiro & Giroux 1987; Donahue & Shull 1991)
- Compton heating from UV photons
- Mechanical heating from supernova-driven winds (Schwartz *et al.* 1975; Ikeuchi 1981; Ostriker & Cowie 1981)
- Cosmic ray heating (Ginzburg & Ozernoi 1965)

The second effect tends to drive the IGM temperature towards two-thirds of the temperature of the ionizing radiation, whereas the first effect tends to drive the temperature towards a lower value T^* that will be defined below. The third and fourth effect can produce much higher temperatures, often in the millions of degrees. The higher the temperature, the greater the y-distortion.

In the GO models, the second effect dominates, which is why they fail so badly. In this chapter, we wish to place limits that are virtually impossible to evade. Thus we will use the most cautions assumptions possible, and assume that the latter three heating mechanisms are negligible.

5.2 The Compton y-Parameter

Thomson scattering between CBR photons and hot electrons affects the spectrum of the CBR. It has long been known that hot ionized IGM causes spectral distortions to the CBR, known as the Sunyaev-Zel'dovich effect. A useful measure of this distortion is the Comptonization y-parameter (Kompanéets 1957; Zel'dovich & Sunyaev 1969; Stebbins & Silk 1986; Bartlett & Stebbins 1991)

$$y = \int \left(\frac{kT_e - kT_\gamma}{m_e c^2}\right) n_e \sigma_t c\, dt = y^* \int \frac{(1+z)}{\sqrt{1+\Omega_0 z}} \Delta T_4(z) x(z) dz, \tag{5.1}$$

where

$$y^* \equiv \left[1-\left(1-\frac{x_{He}}{4x}\right)Y\right]\left(\frac{k\times 10^4\mathrm{K}}{m_e c^2}\right)\left(\frac{3H_0\Omega_{igm}\sigma_t c}{8\pi G m_p}\right) \approx 9.58\times 10^8 h\Omega_{igm}.$$

Here T_e is the electron temperature, T_γ is the CBR temperature, $\Delta T_4 \equiv (T_e - T_\gamma)/10^4$K, Ω_{igm} is the fraction of critical density in intergalactic medium, and $x(z)$ is the fraction of the hydrogen that is ionized at redshift z. Note that we may have $\Omega_{igm} \ll \Omega_b$, *i.e.* all baryons may not be in diffuse form. The integral is to be taken from the reionization epoch to today. In estimating the electron density n_e, we have taken the mass fraction of helium to be $Y \approx 24\%$ and assumed $x_{He} \approx x$, *i.e.* that helium never becomes doubly ionized and that the fraction that is singly ionized equals the fraction of hydrogen that is ionized. The latter is a very crude approximation, but makes a difference of only 6%.

Let us estimate this integral by making the approximation that the IGM is cold and neutral until a redshift z_{ion}, at which it suddenly becomes ionized, and after which it remains completely ionized with a constant temperature T_e. Then for $z_{ion} \gg 1$ and $T_e \gg z_{ion}\times 2.7$K we obtain

$$y \approx 6.4\times 10^{-8} h\Omega_{igm}\Omega_0^{-1/2} T_4\, z_{ion}^{3/2},$$

where $T_4 \equiv T_e/10^4$K. Substituting the most recent observational constraint from the COBE FIRAS experiment, $y < 2.5\times 10^{-5}$ (Mather *et al.* 1994), into this expression yields

$$z_{ion} < 554 T_4^{-2/3}\Omega_0^{1/3}\left(\frac{h\Omega_{igm}}{0.03}\right)^{-2/3}. \tag{5.2}$$

Thus the only way to have z_{ion} as high as 1100 is to have temperatures considerably below 10^4K. In the following section, we will see to what extent this is possible.

5.3 IGM Evolution in the Strong UV Flux Limit

In this section, we will calculate the thermal evolution of IGM for which

- the IGM remains almost completely ionized at all times,
- the Compton y-distortion is minimized given this constraint.

5.3.1 The ionization fraction

In a homogeneous IGM at temperature T exposed to a density of ζ UV photons of energy $h\nu > 13.6\,\mathrm{eV}$ per proton, the ionization fraction x evolves as follows:

$$\frac{dx}{d(-z)} = \frac{1+z}{\sqrt{1+\Omega_0 z}}\left[\lambda_{pi}(1-x) + \lambda_{ci}x(1-x) - \lambda_{rec}x^2\right], \tag{5.3}$$

where $H_0^{-1}(1+z)^{-3}$ times the rates per baryon for photoionization, collisional ionization and recombination are given by

$$\begin{cases} \lambda_{pi} \approx 1.04\times 10^{12}\,[h\Omega_{igm}\sigma_{18}]\,\zeta, \\ \lambda_{ci} \approx 2.03\times 10^4 h\Omega_{igm} T_4^{1/2} e^{-15.8/T_4}, \\ \lambda_{rec} \approx 0.717 h\Omega_{igm} T_4^{-1/2}\left[1.808 - 0.5\ln T_4 + 0.187 T_4^{1/3}\right], \end{cases} \tag{5.4}$$

and $T_4 \equiv T_e/10^4$K. Here σ_{18} is the spectrally-averaged photoionization cross section in units of $10^{-18}\mathrm{cm}^2$. The differential cross section is given by (Osterbrock 1974)

$$\frac{d\sigma_{18}}{d\nu}(\nu) \approx \begin{cases} 0 & \text{if } \nu < 13.6\,\mathrm{eV}, \\ 6.30\frac{e^{4-4\arctan(\epsilon)/\epsilon}}{\nu^4\left(1-e^{-2\pi/\epsilon}\right)} & \text{if } \nu \geq 13.6\,\mathrm{eV}, \end{cases} \tag{5.5}$$

where

$$\epsilon \equiv \sqrt{\frac{h\nu}{13.6\,\mathrm{eV}} - 1}.$$

The recombination rate is the total to all hydrogenic levels (Seaton 1959; Spitzer 1968). Recombinations directly to the ground state should be included here, since as will become evident below, the resulting UV photons are outnumbered by the UV photons that keep the IGM photoionized in the first place, and thus can be neglected when determining the equilibrium temperature.

At high redshifts, the ionization and recombination rates greatly exceed the expansion rate of the universe, and the ionization level quickly adjusts to a quasi-static equilibrium value for which the expression in square brackets in equation (5.3) vanishes. In the absence of photoionization, an ionization fraction x close to unity requires $T_e > 15,000$K. Substituting this into equation (5.2) gives consistency with $z_{ion} > 1000$ only if $h\Omega_{igm} < 0.008$, a value clearly inconsistent with the standard nucleosynthesis constraints (Smith *et al.* 1993). Thus any reheating scenario that relies on collisional ionization to keep the IGM ionized at all times may be considered ruled out by the COBE FIRAS data.

However, this does not rule out all ionized universe scenarios, since photoionization can achieve the same ionization history while causing a much smaller y-distortion. The lowest temperatures (and hence the smallest y-distortions) compatible with high ionization will be obtained when the ionizing flux is so strong that $\lambda_{pi} \gg \lambda_{ci}$. In this limit, to a good approximation, equation (5.3) can be replaced by the following simple model for the IGM:

- It is completely ionized ($x = 1$).
- When a neutral hydrogen atom is formed through recombination, it is instantly photoionized again.

Thus the only unknown parameter is the IGM temperature T_e, which determines the recombination rate, which in turn equals the photoionization rate and thus determines the rate of heating.

5.3.2 The spectral parameter T^*

The net effect of a recombination and subsequent photoionization is to remove the kinetic energy $\frac{3}{2}kT$ from the plasma and replace it with the kinetic energy $\frac{3}{2}kT^*$, where T^* is defined by $\frac{3}{2}kT^* \equiv \langle E_{uv}\rangle - 13.6\,\mathrm{eV}$ and $\langle E_{uv}\rangle$ is the average energy of the ionizing photons. Thus the higher the recombination rate, the faster this effect will tend to push the temperature towards T^*.

UV source	Spectrum $P(\nu)$	$\langle E_{uv}\rangle$	T^*
O3 star	$T = 50,000$K Planck	17.3 eV	28,300K
O6 star	$T = 40,000$K Planck	16.6 eV	23,400K
O9 star	$T = 30,000$K Planck	15.9 eV	18,000K
Pop. III star	$T = 50,000$K Vacca	18.4 eV	36,900K
Black hole, QSO	$\alpha = 1$ power law	18.4 eV	37,400K
?	$\alpha = 2$ power law	17.2 eV	27,800K
?	$\alpha = 0$ power law	20.9 eV	56,300K
?	$T = 100,000$K Planck	19.9 eV	49,000K

Table 5.1: Spectral parameters

The average energy of the ionizing photons is given by the spectrum $P(\nu)$ as $\langle E_{uv}\rangle = h\langle\nu\rangle$, where

$$\langle\nu\rangle = \frac{\int_0^\infty P(\nu)\sigma(\nu)d\nu}{\int_0^\infty \nu^{-1}P(\nu)\sigma(\nu)d\nu}.$$

Here σ is given by equation (5.5). Note that, in contrast to certain nebula calculations where all photons get absorbed sooner or later, the spectrum should be weighted by the photoionization cross section. This is because most photons never get absorbed, and all that is relevant is the energy distribution of those photons that do. Also note that $P(\nu)$ is the energy distribution (W/Hz), not the number distribution which is proportional to $P(\nu)/\nu$.

The spectral parameters $\langle E_{uv}\rangle$ and T^* are given in Table 5.1 for some selected spectra (this is merely a subset of Table 3.3). A power law spectrum $P(\nu) \propto \nu^{-\alpha}$ with $\alpha = 1$ fits observed QSO spectra rather well in the vicinity of the Lyman limit (Cheney & Rowan-Robinson 1981; O'Brien *et al.* 1988), and is also consistent with the standard model for black hole accretion. A Planck spectrum $P(\nu) \propto \nu^3/\left(e^{h\nu/kT} - 1\right)$ gives a decent prediction of T^* for stars with surface temperatures below 30,000K. For very hot stars, more realistic spectra (Vacca 1993) fall off much slower above the Lyman limit, thus giving higher values of T^*. As seen in Table 5.1, an extremely metal poor star of surface temperature 50,000K gives roughly the same T^* as QSO radiation. The only stars that are likely to be relevant to early photoionization scenarios are extremely hot and short-lived ones, since the universe is less than a million years old at $z = 1000$, and fainter stars would be unable to inject enough energy in so short a time. Conceivably, less massive stars could play a the dominant role later on, thus lowering T^*. However, since they radiate such a small fraction of their energy above the Lyman limit, very large numbers would be needed, which could be difficult to reconcile with the absence of observations of Population III stars today.

5.3.3 The thermal evolution

At the low temperatures involved, the two dominant cooling effects[1] are Compton drag against the microwave background photons and cooling due to the adiabatic expansion of

[1] Another cooling mechanism is collisional excitation of atomic hydrogen followed by radiative de-excitation, which cools the IGM at a rate of (Dalgarno & McCray 1972)

$$\mathcal{H}_{ce} \approx 7.5\times10^{-19}e^{-11.8/T_4}n^2(1-x)x\,\mathrm{erg\,cm^{-3}}s^{-1}.$$

Model	Ω_0	Ω_{igm}	h	T^*	z_{ion}	$y/0.000025$	Verdict
QSO BDM I	0.15	0.15	0.8	37,400K	1100	6.30	Ruled out
QSO BDM II	0.15	0.15	0.8	37,400K	200	1.43	Ruled out
O9 BDM	0.15	0.15	0.8	18,000K	800	2.91	Ruled out
QSO BDM III	0.15	0.04	0.8	37,400K	1100	0.67	OK
QSO CDM I	1	0.06	0.5	37,400K	1100	0.17	OK
QSO CDM II	1	0.03	0.8	37,400K	1100	0.16	OK

Table 5.2: Compton y-parameters for various scenarios

the universe. Combining these effects, we obtain the following equation for the thermal evolution of the IGM:

$$\frac{dT}{d(-z)} = -\frac{2}{1+z}T + \frac{1+z}{\sqrt{1+\Omega_0 z}}\left[\lambda_{comp}(T_\gamma - T) + \frac{1}{2}\lambda_{rec}(T)(T^* - T)\right], \tag{5.6}$$

where

$$\lambda_{comp} = \frac{4\pi^2}{45}\left(\frac{kT_\gamma}{\hbar c}\right)^4 \frac{\hbar\sigma_t}{H_0 m_e}(1+z)^{-3} \approx 0.00417h^{-1}(1+z)$$

is $(1+z)^{-3}$ times the Compton cooling rate per Hubble time and $T_\gamma = T_{\gamma 0}(1+z)$. The factor of $\frac{1}{2}$ in front of λ_{rec} is due to the fact that the photoelectrons end up sharing their energy with the protons. We have taken $T_{\gamma 0} \approx 2.726$K (Mather *et al.* 1994). Numerical solutions to this equation are shown in Figure 5.1, and the resulting y-parameters are given in Table 5.2.

The temperature evolution separates into two distinct phases. In the first phase, which is almost instantaneous due to the high recombination rates at low temperatures, T rises very rapidly, up to a quasi-equilibrium temperature slightly above the temperature of the microwave background photons. After this, in the second phase, T changes only slowly, and is approximately given by setting the expression in square brackets in equation (5.6) equal to zero. This quasi-equilibrium temperature is typically much lower than T^*, since Compton cooling is so efficient at the high redshifts involved, and is given by

$$\Delta T \equiv T_e - T_\gamma \approx \frac{\lambda_{rec}}{2\lambda_{comp}}(T^* - T_e) \propto \frac{1}{1+z}g(T_e)h^2\Omega_{igm}(T^* - T), \tag{5.7}$$

independent of Ω_0, where $g(T_e) \gtrsim T^{-0.7}$ encompasses the temperature dependence of λ_{rec}. We typically have $T \ll T^*$. Using this, making the crude approximation of neglecting the temperature dependence of λ_{rec}, and substituting equation (5.7) into equation (5.1) indicates that

$$y \gtrsim h^3\Omega_{igm}^2\Omega_0^{-1/2}T_4^* z_{ion}^{1/2}.$$

The ratio of this cooling rate to the Compton cooling rate is

$$\frac{\mathcal{H}_{ce}}{\mathcal{H}_{comp}} \approx \frac{\exp\left[9.2 - 11.8/T_4\right](1-x)}{(1+z)T_4}h^2\Omega_{igm},$$

a quantity which is much smaller than unity for any reasonable parameter values when $T < 10^4$K. As will be seen, the temperatures at $z \approx 1000$ are typically a few thousand K, which with $h^2\Omega_{igm} < 0.1$ and $x > 0.9$ renders collisional excitation cooling more than nine orders of magnitude weaker than Compton cooling. Hence we can safely neglect collisional excitations when computing the IGM temperature, the reason essentially being that the temperatures are so low that this process is suppressed by a huge Boltzman factor.

Numerically selecting the best power-law fit, we find that this is indeed not too far off: the approximation

$$y \approx 0.0012 h^{2.4} \Omega_{igm}^{1.8} \Omega_0^{-1/2} (T_4^*)^{0.8} (z_{ion}/1100)^{0.9} \tag{5.8}$$

is accurate to about 10% within the parameter range of cosmological interest. We have used equation (5.8) in Figure 5.3 by setting $y = 2.5 \times 10^{-5}$ and $z_{ion} = 1100$. The shaded region of parameter space is thus ruled out by the COBE FIRAS experiment for fully ionized scenarios.

5.4 Conclusions

A reanalysis of the Compton y-distortion arising from early reionization shows that despite the radical sharpening of the FIRAS limit on y, one still cannot conclusively assert that the universe underwent a neutral phase. Non-recombining scenarios where the ionization is thermal, caused by IGM temperatures exceeding 10^4K, are clearly ruled out. Rather, the loophole is for the dominant ionization mechanism to be photoionization. We have shown that for spectra characteristic of both QSO radiation and massive metal-poor stars, the resulting IGM temperatures are so low that typical CDM models with no recombination can still survive the FIRAS test by a factor of six. This conclusion is valid if the flux of ionizing radiation is not so extreme that Compton heating becomes important. This is not difficult to arrange, as the cross section for Thomson scattering is some six orders of magnitude smaller than that for photoionization.

For BDM models, the constraints are sharper. Non-recombining "classical" BDM models with $\Omega_{igm} = \Omega_0 \approx 0.15$ are ruled out even with the extremely cautious reheating assumptions used in this chapter, the earliest ionization redshift allowed being $z \approx 130$. Such models involving early non-linear seeds that on energetic grounds can very plausibly provide a photoionization source capable of reionizing the universe soon after the period of first recombination inevitably generate Compton distortions of order 10^{-4}. These include texture as well as BDM models, both of which postulate, and indeed require, early reionization ($z > 100$) to avoid the generation of excessive anisotropy in the cosmic microwave background on degree angular scales.

Thus BDM models with reionization at $z \approx 1000$ can survive the y test only if most of the baryons form BDM when reionization occurs, and are thereby removed as a source of y-distortion, at least in the diffuse phase. This may be difficult to arrange at $z > 100$, since once the matter is reionized at this high a redshift, Compton drag is extremely effective in inhibiting any further gas collapse until $z < 100$. Since it takes only a small fraction of the baryons in the universe to provide a source of photons sufficient to maintain a fully ionized IGM even at $z \sim 1000$, we suspect that most of the baryons remain diffuse until Compton drag eventually becomes ineffective. Moreover, the possibility that the IGM is only partially reionized at $z \sim 1000$ (*e.g.* GO), a situation which allows a lower value of the y-parameter, seems to us to be implausible as a delicate adjustment of ionization and recombination time-scales over a considerable range in z would be required. A complementary argument that greatly restricts the parameter space allowable for fully ionized BDM models appeals to temperature fluctuations induced on the secondary last scattering surface, both by first order Doppler terms on degree scales and by second order terms on subarcminute scales (Hu *et al.* 1994). Thus, BDM models would seem to be in some difficulty because of the low limit on a possible y-distortion.

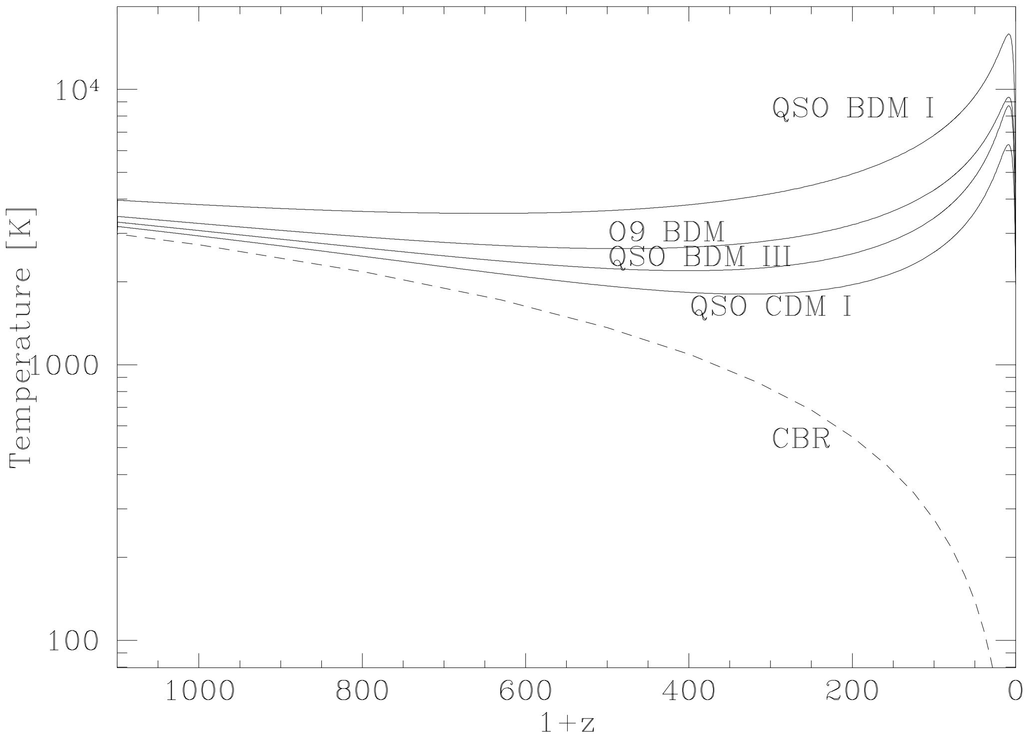


Figure 5.1: Thermal histories for various models.

The temperature of the photoionized IGM is plotted for four of the cosmological models and spectra of ionizing radiation listed in Table 5.2. The lowermost curve gives the temperature of the CBR photons.

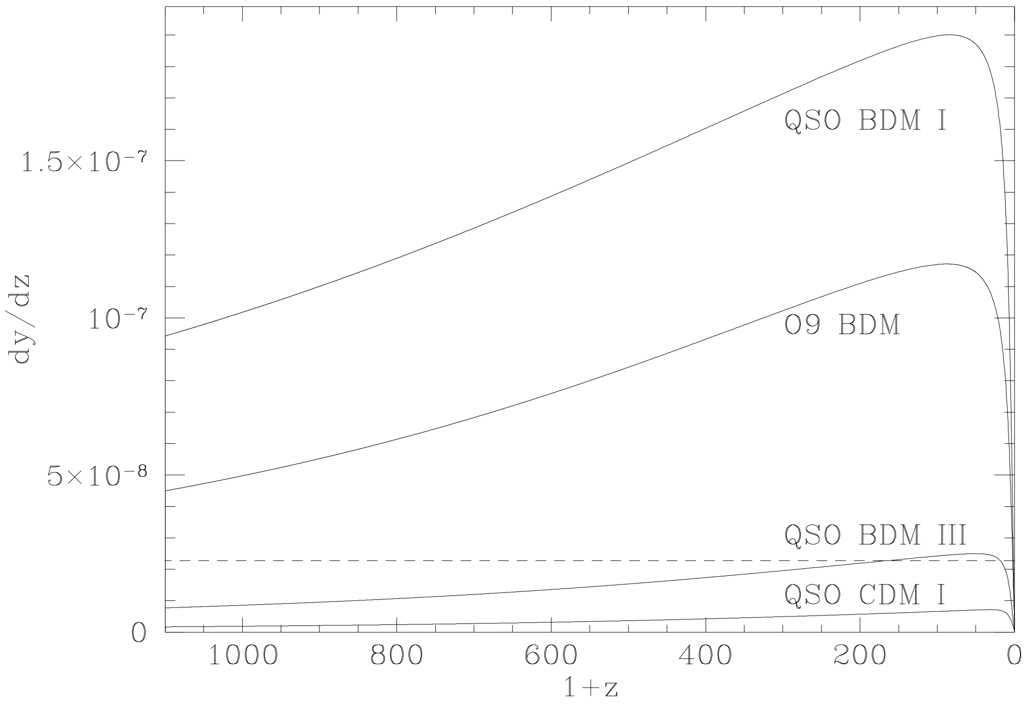


Figure 5.2: dy/dz for various models

The contribution to the y-parameter from different redshifts is plotted four of the cosmological models and spectra of ionizing radiation listed in Table 5.2. Thus for each model, the area under the curve is the predicted y-parameter. The area under the horizontal dashed line is 2.5×10^{-5}, *i.e.* the COBE FIRAS limit.

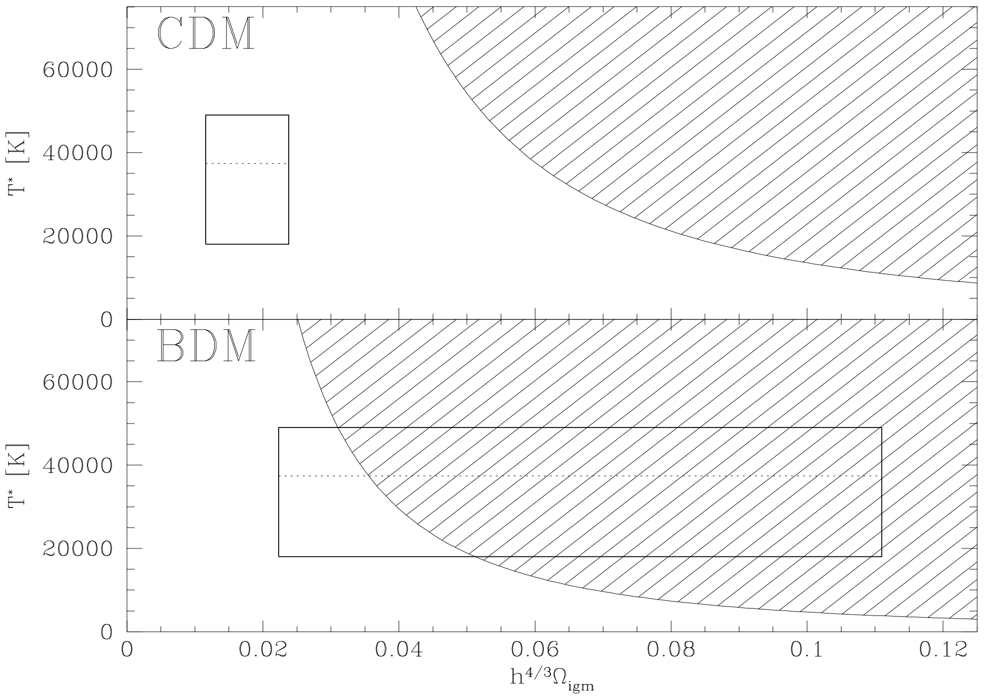


Figure 5.3: Predicted and ruled out regions of parameter space.

The hatched regions of parameter space are ruled out by the the COBE FIRAS limit $y < 2.5 \times 10^{-5}$ for $z_{ion} = 1100$. $\Omega_0 = 1$ in the CDM plot and $\Omega_0 = 0.15$ in the BDM plot. The rectangular regions are the assumed parameter values for the CDM and BDM models, respectively. For CDM, the range $0.012 < h^{4/3}\Omega_{igm} < 0.024$ is given by the nucleosynthesis constraint $0.010 < h^2\Omega_b < 0.015$ and the assumption that $0.5 < h < 0.8$. (If $\Omega_{igm} < \Omega_b$, the rectangle shifts to the left.) For the BDM models, $h = 0.8$ and $0.03 \leq \Omega_{igm} \leq \Omega_0$. The vertical range corresponds to feasible values of the spectral parameter T^*. The upper limit corresponds to highly speculative star with surface temperature 100,000K and $T^* = 49,000$K. The lower line corresponds to an O9 star. The dotted horizontal line corresponds to the spectrum expected from quasars/accreting black holes.

Chapter 6

Late Reionization by Supernova-Driven Winds

In this chapter, a model is presented in which supernova-driven winds from early galaxies reionize the intergalactic medium by $z = 5$. This scenario can explain the observed absence of a Gunn-Peterson trough in the spectra of high-redshift quasars providing that the bulk of these early galaxies are quite small, no more massive than about $10^8 M_\odot$. It also predicts that most of the IGM was enriched to at least 10% of current metal content by $z = 5$ and perhaps as early as $z = 15$. The existence of such early mini-galaxies violates no spectral constraints and is consistent with a pure CDM model with $b \leq 2$. Since the final radius of a typical ionized bubble is only around 100 kpc, the induced modification of the galaxy autocorrelation function is negligible, as is the induced angular smoothing of the CBR. Some of the gas swept up by shells may be observable as pressure-supported Lyman-alpha forest clouds.

6.1 Introduction

The absence of a Gunn-Peterson trough in the spectra of high-redshift quasars has provided strong evidence for the intergalactic medium (IGM) being highly ionized as early as $z = 4$ (Gunn & Peterson 1965; Steidel and Sargent 1987; Webb *et al.* 1992). The hypothesis that photoionization of the IGM by quasars could account for this ionization (Arons & McCray 1969; Bergeron & Salpeter 1970; Sherman 1980) has been challenged (Shapiro 1986; Shapiro & Giroux 1987; Miralda-Escude & Ostriker 1990). Other studies have maintained that photoionization by quasars (Donahue & Shull 1987) or active galactic nuclei (Teresawa 1992) may nonetheless be sufficient. However, in view of the large uncertainties in crucial parameters such as ionizing fluxes, the issue of what reionized the IGM must still be considered open.

In comparing the Gunn-Peterson constraints with our work in Chapters 3 and 4, the crucial difference is the degree of ionization required. To affect the CBR, it does not really matter whether the ionization fraction x is 90% or 99.999%, as this makes only a 10% difference in the optical depth τ. The Gunn-Peterson limits constrain not x but $(1 - x)$, the neutral fraction. Thus in this context, the difference between 90% and 99.999% ionization is four orders of magnitude. In Chapters 3 and 4, we found that photoionization by early galaxies could easily reionize the IGM by a redshift $z = 5$, but the issue here is whether photoionization alone can provide the extremely high ionization fraction required to pass the Gunn- Peterson test.

In this chapter, we investigate an alternative reionization scenario, which produces considerably higher IGM temperatures than those attained by the photoionization models in previous chapters. Supernova driven winds from luminous galaxies have long been conjectured to be an important ionization source for the IGM (Schwartz *et al.* 1975; Ikeuchi & Ostriker 1986; Carlberg & Couchman 1989). Cold dark matter (CDM)-based models of structure formation (Blumenthal *et al.* 1984; Efstathiou *et al.* 1985) predict the formation of gravitationally bound objects of mass as small as $10^7 M_\odot$ in large numbers before $z = 5$. Recent work (Blanchard *et al.* 1992) indicates that such objects can cool rapidly and presumably fragment into stars. These early mini-galaxies would be expected to release great amounts of kinetic energy into the surrounding IGM, thereby creating large, fairly spherical voids filled with thin, hot, ionized plasma. We analyze the effect of expanding bubbles driven by supernova winds from early mini-galaxies, and show that this mechanism of distributing energy can indeed provide the required ionization without violating any of the current spectral constraints.

In Section 6.2, we will treat the expansion of a shell in a uniform, cold and neutral IGM. As these bubbles become larger and more numerous and fill most of space, this obviously becomes a very poor model of the IGM. In Section 6.3 we estimate bulk properties of this new processed IGM such as temperature, density and ionization.

6.2 The Explosion Model

Since the pioneering work on spherically symmetric explosions by Sedov (1959), a profusion of analytic solutions have been given by numerous authors for models of ever-increasing complexity (Cox & Smith 1974; McKee & Ostriker 1977; Weaver *et al.* 1977; McCray & Snow 1979; Bruhweiler *et al.* 1980; Tomisaka *et al.* 1980; McCray & Kafatos 1987; Ostriker & McKee 1988). Most of these models pertain to bubbles in the interstellar medium of a galaxy, where the expansion of the universe can be ignored. Ostriker & McKee have given asymptotic self-similarity solutions that incorporate this latter complication, but unfortunately they are not sufficiently accurate for our needs. The reason is that since neither energy nor momentum is conserved in the regime before the shell becomes self-similar, there is no accurate way to normalize the self-similar solution using the initial data.

Let ρ_b and ρ_d denote the average densities of baryonic and non-baryonic matter in the universe. We will assume that all baryons are in diffuse for early on, so that ρ_b is also the density of the IGM. We will write $\rho_b = \Omega_b \rho_c$ and $\rho_d = \Omega_d \rho_c$, where the critical density $\rho_c \equiv 3H^2/8\pi G$. We will use a three-phase model for the expanding bubbles:

- a dense, fairly cool spherical shell of outer radius R and thickness $R\delta$, containing a fraction $(1 - f_m)$ of the total baryonic mass enclosed,

- uniform neutral ambient intergalactic medium (IGM) of density $\rho_b + \rho_d$ and zero pressure outside,
- a hot, thin, isothermal plasma of pressure p and temperature T inside the shell.

The shell is driven outwards by the pressure of the hot interior plasma but slowed by the IGM and by gravity. The plasma is heated by kinetic energy from supernova explosions and collisions with IGM and cooled by bremsstrahlung and Compton drag against the cosmic background radiation.

6.2.1 The expanding shell

We assume that the expanding shell sweeps up almost all baryonic IGM that it encounters and loses only a small fraction of it through evaporation into the interior, so that its total mass is given by $m(t) = \frac{4}{3}\pi R(t)^3(1-f_m)\rho_b$, where the constant $f_m \ll 1$. Since $\dot{\rho_b}/\rho_b = -3H$ for any cosmological model, we get

$$\frac{\dot{m}}{m} = \left(R^3\rho_b\right)^{-1}\frac{d}{dt}\left(R^3\rho_b\right) = 3\left(\frac{\dot{R}}{R} - H\right) \text{ if } \frac{\dot{R}}{R} > H, \text{ zero otherwise.}$$

(The shell will acquire new mass when it is expanding faster than the Hubble flow, and will never lose mass.) It turns out that the Hubble flow catches up with the shell only as $t \to \infty$, so we will always have $\dot{R} > HR$ and $\dot{m} > 0$.

When new mass is swept up, it must be accelerated from the velocity HR to $\dot{R}$, so the shell experiences a net braking force $(\dot{R} - HR)\dot{m}$. The interior pressure p drives the shell outward with a force $pA = 4\pi R^2 p = 3mp/\rho_b R$ in the thin shell approximation $\delta, f_m \ll 1$. Finally there is a gravitational braking force, which in the thin-shell approximation (Ostriker & McKee 1988) gives the deceleration $\frac{4}{3}\pi GR(\rho_d + \frac{1}{2}\rho_b)$. Adding these three force terms, the radial equation of motion becomes

$$\ddot{R} = \frac{8\pi pG}{\Omega_b H^2 R} - \frac{3}{R}\left(\dot{R} - HR\right)^2 - \left(\Omega_d + \frac{1}{2}\Omega_b\right)\frac{H^2 R}{2}. \tag{6.1}$$

6.2.2 The interior plasma

The equation of state for the plasma in the hot interior gives the thermal energy

$$E_t = \frac{3}{2}pV = 2\pi pR^3, \tag{6.2}$$

and energy conservation for the interior yields

$$\dot{E}_t = L - pdV/dt = L - 4\pi pR^2\dot{R}, \tag{6.3}$$

where the luminosity L incorporates all sources of heating and cooling of the plasma. We will consider five contributions to L and write

$$L = L_{sn} - L_{comp} - L_{brems} - L_{ion} + L_{diss},$$

where L_{sn} is the energy injection from supernova explosions, L_{comp} the cooling by Compton drag against the CBR, L_{brems} the cooling by bremsstrahlung, L_{ion} the cooling by ionization of neutral IGM and L_{diss} the heating from collisions between the shell and the IGM.

In stellar burning from zero to solar metallicity, the mass fraction 0.02×0.007 is released, mostly as radiation. Due to low cross-sections, only a negligible fraction of this radiation will contribute towards heating the gas, so we will only be interested in the energy that is released in kinetic form. From empirical observations of active galactic winds (Heckman 1990) about 2% of the total luminosity from a galaxy is mechanical. Another empirical observation is that for a solar neighborhood initial stellar mass function, one has roughly one supernova for every $150 M_\odot$ of baryons that form stars, with a typical kinetic energy output of 10^{51} ergs per explosion. Both of these observations lead to the same estimate

$$L_{sn} = \frac{f_{sn} M_b c^2}{t_{burn}} \theta(t_{burn} - t) \approx 1.2 L_\odot \frac{M_b}{M_\odot} \theta(t_{burn} - t),$$

where the efficiency $f_{sn} \approx 4 \times 10^{-6}$ and where we have assumed that the energy is released at a constant rate during a period $t_{burn} \approx 5 \times 10^7$ years.

Now let us examine cooling. The interior baryon density is $\rho_i = \rho_b f_m / (1 - \delta)^3$ whereas the shell density is $\rho_s = \rho_b (1 - f_m) / (1 - (1 - \delta)^3) \approx \rho_b / 3\delta$ if $f_m, \delta \ll 1$. Compton drag against the microwave background radiation causes energy loss at a rate (Kompaneets 1957)

$$L_{comp} = \frac{4\pi^2}{15} (\sigma_t c n_e) \left(\frac{kT_e}{m_e c^2} \right) \left(\frac{kT_\gamma}{\hbar c} \right)^4 \hbar c V, \tag{6.4}$$

where σ_t is the Thomson cross section, $V = \frac{4}{3}\pi R^3$, and $T_e = T$, the temperature of the interior plasma, which is given by

$$E_t = \left(\frac{3}{2} + \frac{3}{2} \right) kT \frac{f_m \rho_b}{m_p} V.$$

(We will assume almost complete ionization and low metallicity, so that $n_e \approx f_m n_b$.) Using equation (6.2), we see that Compton drag causes cooling on a timescale

$$\frac{E_t}{L_{comp}} = \frac{45}{4\pi^2} \left(\frac{\hbar c}{kT_\gamma} \right)^4 \frac{m_e}{\sigma_t \hbar} \approx 2 \times 10^{12} \,\text{years} \times (1 + z)^{-4},$$

that is, it becomes important only at high redshifts. It turns out that $L_{brems} \ll L_{comp}$ in our regime of interest, so we will simply make the approximation $L_{brems} \approx 0$. Assuming that the ambient IGM is completely neutral, the power required to ionize the hydrogen entering the interior is simply

$$L_{ion} = f_m n_b E_0 \times 4\pi R^2 \left[\dot{R} - HR \right],$$

where $E_0 \approx 13.6\, eV$.

The equation of motion (6.1) assumes that the collisions between the expanding shell and the ambient IGM are perfectly inelastic. The kinetic energy dissipated has one of three fates: It may

(a) radiate away in shock cooling processes,

(b) ionize the swept up IGM, or

(c) heat the shell and by conduction heat the interior plasma.

Let f_d denote the fraction that is reinjected into the interior plasma through processes (b) and (c). This is one of the major uncertainties of the model. Now a straightforward kinematic calculation of the kinetic energy loss per unit time gives

$$L_{diss} = f_d \frac{3m}{2R}\left(\dot{R} - HR\right)^3.$$

Making accurate estimates of f_d is difficult, so we simply use the two extreme cases $f_d = 0$ and $f_d = 1$ in the simulations. Perhaps surprisingly, the results will be seen to be relatively independent of the choice of f_d.

6.2.3 Solutions to the equations

Combining (6.2) and (6.3) leaves us with

$$\dot{p} = \frac{L}{2\pi R^3} - 5\frac{\dot{R}}{R}p. \tag{6.5}$$

The system of equations (6.1) and (6.5) reduces to that derived by Weaver *et al.* (1977) in the special case where $L(t)$ is constant and the expansion of the universe is ignored.

Let us define dimensionless variables as follows:

$$\begin{array}{llll}
\tau \equiv t/t_*, & t_* \equiv & \frac{2}{3}H_0^{-1}(1+z_*)^{-3/2} & \\
\eta \equiv H/H_*, & H_* \equiv & \frac{2}{3}t_*^{-1} & \\
\ell \equiv L/L_*, & L_* \equiv & f_{sn}M_b c^2 t_{burn}^{-1} & \approx 1.2\times 10^5 L_\odot \times M_5 \\
\varepsilon \equiv E/E_*, & E_* \equiv & L_* t_{burn} & \approx 7.2\times 10^{53}\,\mathrm{erg}\times M_5 \\
r \equiv R/R_*, & R_* \equiv & L_*^{1/5}G^{1/5}t_* & \approx 0.13\,\mathrm{Mpc}\times h^{-1}(1+z_*)^{-3/2}M_5^{1/5} \\
q \equiv p/p_*, & p_* \equiv & L_*^{2/5}G^{-3/5}t_*^{-2} & \approx 1.4\times 10^{-16}\,Pa\times h^2(1+z_*)^3 M_5^{2/5}
\end{array}$$

Here we have taken $h = 0.5$ and defined $M_5 \equiv M_b/10^5\,M_\odot$. If $\Omega \equiv \Omega_b + \Omega_d = 1$, then t_* is the age of the universe at the redshift z_* when the shell begins its expansion, i.e. the Big Bang occurred at $\tau = -1$ and the shell starts expanding at $\tau = 0$. For this simple case, we have $\eta = (1+\tau)^{-1} = (1+z_*)^{-3/2}(1+z)^{3/2}$.

Equations (6.1) and (6.5) now become

$$\left\{\begin{array}{ll}
r''(\tau) = & \frac{18\pi}{\Omega_b}\eta(\tau)^{-2}\frac{q(\tau)}{r(\tau)} - 3\left(1 - \frac{2}{3}\frac{\eta(\tau)r(\tau)}{r'(\tau)}\right)^2\frac{r'(\tau)^2}{r(\tau)} - \left(\frac{2}{9}\Omega_d + \frac{1}{9}\Omega_b\right)\eta(\tau)^2 r(\tau) \\
q'(\tau) = & \frac{\ell(\tau)}{2\pi r(\tau)^3} - 5\frac{r'(\tau)}{r(\tau)}q(\tau)
\end{array}\right. \tag{6.6}$$

Here $\ell = \ell_{sn} - \ell_{comp} - \ell_{ion} + \ell_{diss}$, where

$$\begin{array}{lll}
\ell_{sn} & = & \theta(t_{burn} - t_*\tau), \\
\ell_{comp} & \approx & 0.017h^{-1}(1+z_*)^{-3/2}(1+z)^4 r^3 q, \\
\ell_{ion} & \approx & 2.2 f_m \Omega_b M_5^{-2/5} \times \eta^2 r^2 \left(r' - \frac{2}{3}\eta r\right), \text{ and} \\
\ell_{diss} & = & \frac{1}{3} f_d \Omega_b \times (\eta r)^2 \left(r' - \frac{2}{3}\eta r\right)^3.
\end{array}$$

In computing ℓ_{comp}, we have taken $T_{\gamma 0} = 2.74K$. The interior temperature, the thermal energy and the kinetic energy are given by

$$T \approx 4.5 \times 10^5\, K \times \frac{M_5^{2/5}}{f_m \Omega_b} \frac{q(\tau)}{\eta(\tau)^2},$$

$$\varepsilon_t = \frac{2\pi}{\tau_{burn}} r^3 q,$$

$$\varepsilon_k = \frac{m\dot{R}^2/2}{L_* t_{burn}} = \frac{\Omega_b}{9\tau_{burn}} \eta^2 r^3 r'^2.$$

The solution to the system (6.6) evolves through three qualitatively different regimes: $\tau \ll 1$, $\tau \approx 1$ and $\tau \gg 1$.

a) In the limit of small times $\tau \ll 1$, gravity and Hubble flow are negligible and we obtain the asymptotic power law solution

$$r(\tau) = a\tau^{3/5}, \quad q(\tau) = b\tau^{-4/5}$$

$$\text{where } a \equiv \left(\frac{375/\Omega_b}{77 - 27 f_d}\right)^{1/5} \quad \text{and} \quad b \equiv \frac{7\Omega_b}{150\pi} a^2, \tag{6.7}$$

as may be verified by direct substitution. This solution reduces to that found by Weaver *et al.* in the special case $f_d = 0$. Since the total energy injected is simply $\varepsilon_{in} = \tau/\tau_{burn}$, this gives

$$\frac{\varepsilon_t}{\varepsilon_{in}} = \frac{35}{77 - 27 f_d} \quad \text{and} \quad \frac{\varepsilon_k}{\varepsilon_{in}} = \frac{15}{77 - 27 f_d}$$

for small τ. Hence even though $\varepsilon_t + \varepsilon_k = \varepsilon_{in}$ only for the most optimistic case $f_d = 1$, we see that no more than $1 - \frac{35+15}{77} \approx 25\%$ of the injected energy is lost as radiation even in the worst case $f_d = 0$.

b) The behavior in the intermediate regime is a complicated interplay between several different effects:

1. After approximately 5×10^7 years, the supernova explosions cease, which slows the expansion. In this pressure-driven snowplow phase, we would asymptotically have $R \propto t^{2/7}$, $t^{4/13}$, $t^{1/3}$, $t^{4/11}$ or $t^{2/5}$ if there were no gravity, no Hubble flow and no cooling with $f_d = 0$, $\frac{1}{8}$, $\frac{1}{3}$, $\frac{5}{8}$ or 1, respectively.
2. Cooling (and pdV) work reduces the pressure and the thermal energy to virtually zero, which slows the expansion. With zero pressure, we would approach the momentum-conserving snowplow solution $R \propto t^{1/4}$ if there were no gravity and no Hubble flow.
3. The density of the IGM drops and the IGM already has an outward Hubble velocity before it gets swept up, which boosts the expansion and adds kinetic energy to the shell.
4. Gravity slows the expansion.

5. Dark matter that has been accelerated outward by the shell catches up with it again and speeds up the expansion. (This last effect has been neglected in the equations above, since it generally happens too late to be of importance for our purposes.)

c) As $t \to \infty$, the shell gets frozen into the Hubble flow, *i.e.* $R \propto t^{2/3}$ if $\Omega = 1$. An approximate analytic solution for $\tau \gg 1$ is given by Ostriker& McKee (1988), but since neither energy nor momentum is conserved in the intermediate regime, there is no simple way to connect this solution with the short-time solution above.

Numerical solutions for the comoving radius $(1+z)R$ are plotted in Figure 6.1 for different values of z_* and f_d. The asymptotic solution (6.7) has been used to generate initial data at $\tau = 0.01$ for the numerical integration. In this Figure, we have truncated R when the interior temperature drops below 15,000K, after which newly swept up IGM fails to become ionized. Figures 6.2, 6.3 and 6.4 show what becomes of the injected energy for different parameter values. Note that the relative fractions are approximately constant early on, while the supernovae inject energy, in accordance with the asymptotic solution (6.7). The reason that the total energy exceeds 100% of the input is that the shell gobbles up kinetic energy from swept-up IGM that already has an outward Hubble velocity.

6.3 Cosmological Consequences

Once the expanding bubbles discussed in the previous section have penetrated most of space, the IGM will presumably have a frothy character on scales of a few 100 kpc, containing thick and fairly cool shell fragments separated by large, hot, thin and ionized regions that used to be bubble interiors.

In Section 6.3.1, we calculate at what point the IGM becomes frothy, more specifically what fraction of space is covered by expanding shells at each z. In 6.3.2 we discuss the resulting enrichment of the IGM with heavy elements. In 6.3.3 the thermal history of the IGM after this epoch is treated. Finally, in 6.3.4 the residual ionization is computed, given this thermal history, and we discuss the circumstances under which the Gunn-Peterson constraint is satisfied.

6.3.1 IGM porosity

Assuming the standard PS theory of structure formation (Press & Schechter 1974), the fraction of all mass that has formed gravitationally bound objects of total (baryonic and non-baryonic) mass greater than M at redshift z is

$$1 - \mathrm{erf}\left[\frac{\delta_c}{\sqrt{2}\sigma(M)}\right],$$

where $\mathrm{erf}(x) \equiv 2\pi^{-1/2}\int_0^x e^{-u^2}du$ and σ^2 is the linearly extrapolated r.m.s. mass fluctuation in a sphere of radius r_0. The latter is given by top-hat filtering of the power spectrum as

$$\sigma^2 \equiv \left(\frac{\sigma_0}{1+z}\right)^2 \propto \frac{1}{(1+z)^2}\int_0^\infty \left[\frac{\sin kr_0}{(kr_0)^3} - \frac{\cos kr_0}{(kr_0)^2}\right]^2 P(k)dk, \qquad (6.8)$$

where r_0 is given by $\frac{4}{3}\pi r_0^3\rho = M$ and where $P(k)$ is the power spectrum. Although this approach has been criticized as too simplistic, numerical simulations (Efstathiou *et al.* 1988; Efstathiou & Rees 1988; Carlberg & Couchman 1989) have shown that it describes the mass distribution of newly formed structures remarkably well. Making the standard assumption of a Gaussian density field, Blanchard *et al.* (1992) have argued that it is an accurate description at least in the low mass limit. Since we are interested only in extremely low masses such as $10^6 M_\odot$, it appears to suffice for our purposes.

We choose $\delta_c = 1.69$, which is the linearly extrapolated overdensity at which a spherically symmetric perturbation has collapsed into a virialized object (Gott & Rees 1975). Letting f_g denote the fraction of all baryons in galaxies of mass greater than M at z, this would imply that

$$f_g \approx 1 - \operatorname{erf}\left[\frac{1.69(1+z)}{\sqrt{2}\sigma_0(M)}\right] \tag{6.9}$$

if no other forces than gravity were at work. However, it is commonly believed that galaxies correspond only to such objects that are able to cool (and fragment into stars) in a dynamical time or a Hubble time (Binney 1977; Rees & Ostriker 1977; Silk 1977; White & Rees 1978). Hence the above value of f_g should be interpreted only as an upper limit.

A common assumption is that the first galaxies to form have a total (baryonic and dark) mass $M_c \approx 10^6 M_\odot$, roughly the Jeans mass at recombination. Blanchard *et al.* (1992) examine the interplay between cooling and gravitational collapse in considerable detail, and conclude that the first galaxies to form have masses in the range $10^7 M_\odot$ to $10^8 M_\odot$, their redshift distribution still being given by equation (6.9). To keep things simple we will assume that all early galaxies have the same mass M_c and compare the results for $M_c = 2\times 10^6 M_\odot, 10^8 M_\odot$ and $10^{11} M_\odot$.

Let $R(z; z_*)$ denote the radius of a shell at z that was created at z_* by a galaxy of baryonic mass $M_b = \Omega_b M_c$ as in Section 6.2. Then the *naive filling factor*, the total bubble volume per unit volume of the universe, is

$$\phi(z) = \int_z^\infty \frac{4}{3}\pi R(z;z_*)^3 \frac{\rho_b}{M_b}\frac{df_g(z_*)}{d(-z_*)}\,dz_* = \phi_*(1+z)^3 \int_z^\infty \frac{r(z;z_*)^3}{(1+z_*)^{9/2}}\frac{df_g(z_*)}{d(-z_*)}\,dz_*, \tag{6.10}$$

where

$$\phi_* \approx 1600 h^{-1} M_5^{-2/5}(\Omega_b/0.06).$$

Clearly nothing prohibits ϕ from exceeding unity. This means that nearby shells have encountered each other and that certain volumes are being counted more than once. If the locations of the bubbles are uncorrelated, then the fraction of the universe that will be in a bubble, the *porosity*, is given by

$$P \equiv 1 - e^{-\phi}.$$

If the early galaxies are clustered rather than Poisson-distributed, this value is an overestimate. For an extreme (and very unrealistic) example, if they would always come in clusters of size n and the clusters would be much smaller than the typical bubble size of 100 kpc, then it is easy to see that $P \approx 1 - e^{-\phi/n}$. For more realistic cases, simple analytic

$f_g(5)$	$\sigma_0(M_c)$	$b_{cdm,6}$	$b_{cdm,8}$	$b_{tilted,6}$	$b_{tilted,8}$
1%	3.94	4.8	3.5	2.5	2.0
10%	6.18	3.1	2.2	1.6	1.3
20%	7.92	2.4	1.7	1.2	1.0
50%	15.06	1.3	0.9	0.6	0.5

Table 6.1: Correspondence between various ways of normalizing the power spectrum

expressions for P are generally out of reach. Since we expect the clustering to be quite weak, we will use the Poisson assumption for simplicity.

The uppermost panels of Figures 6.5 and 6.6 contain $P(z)$ for various parameter values, calculated numerically from equation (6.10) using the numerical solutions for $r(z; z_*)$. It is seen that the lower mass in Figure 6.5 ($2 \times 10^6 M_\odot$ versus $10^8 M_\odot$) gives higher filling factors, so that the expanding shells fill almost 100% of space by $z = 5$ for three of the four choices of $f_g(5)$. In 6.6, we see that almost 20% of the baryons must be in galaxies by $z = 5$ to achieve this. The greater efficiency of small galaxies is to be expected, since $\phi_* \propto M^{-2/5}$. Although some parameters still yield the desired $P \approx 100\%$ by $z = 5$ in Figure 6.6, using present-day masses like $M_c = 10^{11} M_\odot$ fails dismally (not plotted) for all choices of the other parameters. Roughly, the largest M_c that works is $10^8 M_\odot$

As can be seen, the dependence on f_d (dashed versus solid lines) is rather weak.

In order to calculate $\sigma_0(M_c)$ from the fluctuations observed on larger scales today, we need detailed knowledge of the power spectrum down to very small scales, something which is fraught with considerable uncertainty. For this reason, we have chosen to label the curves by the more physical parameter $f_g(5)$, the fraction of all baryons that have formed galaxies by $z = 5$. The four sets of curves correspond to fractions of 50%, 20%, 10% and 1%. These percentages should be compared with observational estimates of metallicity, as will be discussed in Section 6.3.2.

The second column of Table 6.1 contains the values $\sigma_0(M_c)$ necessary to obtain various values of $f_g(5)$, calculated by inverting the error function in equation (6.9). The last four columns contain the bias factors necessary to yield this value of $\sigma_0(M_c)$ for two choices of power spectra (CDM and n=0.7 tilted CDM) and two choices of cutoff mass ($M_c = 2 \times 10^6 M_\odot$ and $M_c = 10^8 M_\odot$). Thus $b = \gamma/\sigma_0(M_c)$, where we define γ to be the ratio between σ at M_c and σ at $8h^{-1}\,\mathrm{Mpc} \equiv b^{-1}$. Performing the integral (6.8) numerically with the CDM transfer function given by Bardeen *et al.* 1986 (BBKS), $h = 0.5$, $\Omega = 1$, $\Omega_b \ll 1$ and an $n = 1$ Harrison-Zel'dovich initial spectrum gives $\gamma \approx 19.0$ for $M_c = 2 \times 10^6 M_\odot$ and $\gamma \approx 13.6$ for $M_c = 10^8 M_\odot$. Using the CDM transfer function of Bond and Efstathiou (1984) instead gives $\gamma \approx 18.1$ and $\gamma \approx 13.7$, respectively. The BBKS transfer function is more applicable here since it includes the logarithmic dependence that becomes important for very low masses. The BBKS transfer function with a tilted (n=0.7) primordial spectrum yields the significantly lower values $\gamma \approx 9.71$ and $\gamma \approx 7.93$, respectively.

Basically, Table 6.1 shows that any of our values of $f_g(5)$ become consistent with a feasible bias factor for some choice of power spectrum and cutoff mass.

6.3.2 IGM enrichment

These values of $f_g(5)$ should be compared with observational estimates of metallicity, since if the stars in these early mini-galaxies produce the same fractions of heavy elements as do conventional stars, then these percentages are directly linked to the fraction of currently observed metals that were made before $z = 5$. Some of the enriched shells may be observable as quasar absorption line systems, as intracluster gas, and, indirectly, as in the metallicities of old disk and halo stars.

Observations of iron abundances in intracluster gas by HEAO-1, Exosat and Ginga (*e.g.* Mushotzky 1984; Hughes *et al.* 1988; Edge 1989; Hatsukade 1989) have shown that most clusters have abundances between 25% and 50% of the solar value. Einstein observations have showed the presence of a large variety of other heavy elements in the intracluster gas (Lea *et al.* 1982; Rothenflug *et al.* 1984). Most of this gas and some of these metals are believed to be "primordial", since the gas mass in clusters is typically several times greater than the observed stellar mass in the cluster galaxies (Blumenthal *et al.* 1984; David *et al.* 1990; Arnaud *et al.* 1991).

There are indications that the most of these heavy elements may have been produced as recently as around $z = 2-3$, and that the metallicity in the halo gas of some $z \approx 3$ galaxies inferred from QSO absorption line studies are as low as 0.1% of the solar value (Steidel 1990). However, this and other observations of extremely metal-poor objects (Pettini *et al.* 1990) does not necessarily rule out our scenario, since it is highly uncertain whether all the hydrogen in the swept-up IGM would get thoroughly mixed with the metal-rich supernova ejecta.

6.3.3 IGM temperature

Let $T(z;z_*)$ denote the temperature of the interior of a bubble at z that was created at z_* as in Section 6.2. Then the volume-averaged temperature of the IGM is

$$\begin{aligned} T_{IGM}(z) &\equiv \int_z^\infty \frac{4}{3}\pi R(z;z_*)^3 T(z;z_*) \frac{\rho_b}{M_b}\frac{df_g(z_*)}{d(-z_*)}\,dz_* \\ &= \phi_*(1+z)^3\int_z^\infty \frac{r(z;z_*)^3}{(1+z_*)^{9/2}}\,T(z;z_*)\frac{df_g(z_*)}{d(-z_*)}\,dz_*. \end{aligned} \tag{6.11}$$

When ϕ becomes of order unity, the IGM swept up by the expanding shells is no longer cold, neutral and homogeneous, so the treatment in Section 6.2 breaks down. The resulting temperatures will be underestimated, since less thermal energy needs to be expended on heating and ionization. As can be seen in Figures 6.5 and 6.6, the transition from $\phi \ll 1$, where the treatment in Section 6.2 is valid, to $\phi \gg 1$, where the IGM becomes fairly uniform, is quite rapid. Since T_{IGM} defined above is proportional to the thermal energy per unit volume, energy conservation leads us to assume that T_{IGM} remains fairly constant during this transition and therefore is a good estimate of the bulk IGM temperature immediately afterwards. From this time on, we will approximate the IGM outside the scattered dense and cold shell remnants by a uniform isothermal plasma. Applying equations (6.2), (6.3) and (6.4) to the IGM yields the following equation for its thermal evolution:

$$-\frac{d}{dz}T_5 = -\left[\frac{2}{1+z} + A(1+z)^{3/2}\right]T_5 + \ell_{inj}, \tag{6.12}$$

where the first term encompasses cooling from adiabatic expansion and the second term Compton cooling. $T_5 \equiv T_{IGM}/10^5\,K$, $A \equiv 1.5(t_0/t_{comp}) \approx 0.0042h^{-1}$ and $\ell_{inj} \equiv t_0 L_{inj}/k \times 10^5 K$, where L_{inj} is the power injected into the IGM per proton from all heat sources combined. The Compton cooling term is seen to increase with redshift, equaling the adiabatic term at $z \approx 17h^{0.4} - 1$.

In the most pessimistic case of no reheating whatsoever, *i.e.* for $\ell_{inj} = 0$, equation (6.12) has the solution

$$T \propto (1+z)^2 e^{0.4A(1+z)^{5/2}}.$$

A more optimistic assumption is that some fraction f_{inj} of the total energy released from stellar burning in newly formed galaxies continues to heat the IGM, i.e.

$$\ell_{inj} = f_{inj}\left(\frac{0.02 \times 0.007 m_p c^2}{k \times 10^5 K}\right)\frac{df_g}{d(-z)} \approx \left(\frac{2.1 \times 10^4}{\sigma_0}\right) f_{inj} \exp\left(-\frac{1}{2}\left[\frac{1.69(1+z)}{\sigma_0(M_c)}\right]^2\right),$$

which would also incorporate other modes of energy injection such as radiation. The middle panels in Figures 6.5 and 6.6 show the temperature resulting from $f_{inj} = 0, 0.001$ and 0.02 and different initial values. For $P < 0.8$, T_{IGM} has been calculated numerically from equation (6.11) by using the numerical solutions for $r(z; z_*)$. Then the value at the redshift for which $P = 0.8$ has been used as initial data for equation (6.12). For comparison, two horizontal lines have been added showing what temperatures would be required to obtain neutral fractions of 10^{-6} and 10^{-5} in *equilibrium*, using equation (6.13). Since the plasma is in fact out of equilibrium, these highly ionized states can be maintained at much lower temperatures, as is seen in the bottom plots.

6.3.4 IGM ionization and the Gunn-Peterson effect

Assuming that any neutral hydrogen in the remains of the shells will have insufficient time to diffuse far into the hot ionized regions that used to be shell interiors, we can treat the latter as an isolated mixture of gas and plasma where the ionization fraction χ evolves as

$$\dot{\chi} = n\chi\left[(1-\chi)\Lambda_{ion}(T) - \chi\Lambda_{rec}(T)\right],$$

and where the rates for collisional ionization and recombination are given by (Stebbins & Silk 1986)

$$\Lambda_{ion} = \langle\sigma_{ci}v\rangle \approx 7.2\,a_0^2\left(\frac{kT}{m_e}\right)^{1/2} e^{-E_0/kT},$$

$$\Lambda_{rec} = \langle\sigma_{rec}v\rangle \approx \frac{64\pi}{3\sqrt{3\pi}}\alpha^4 a_0^2 c\left(\frac{kT}{E_0}\right)^{-2/3},$$

where a_0 is the Bohr radius and α is the fine structure constant. Changing the independent variable to redshift, the $\Omega = 1$ case leaves us with

$$-\frac{d}{dz}\chi = \frac{3}{2}\frac{f_m\Omega_b}{(1-\delta)^3}(1+z)^{1/2}\chi[(1-\chi)\lambda_{ion}(\chi) - \chi\lambda_{rec}(T)],$$

$$\lambda_{ion}(T) \equiv n_{c0} t_0 \Lambda_{ion}(T) \approx \left[5.7 \times 10^4 h\Omega_b\right] T_5^{1/2} e^{1.58/T_5}$$

$$\lambda_{rec}(T) \equiv n_{c0} t_0 \Lambda_{rec}(T) \approx [0.16 h\Omega_b]\, T_5^{-2/3}.$$

For large enough z, the ionization fraction will adjust rapidly enough to remain in a quasistatic equilibrium and hence be given by $\dot{\chi} = 0$, *i.e.*

$$\chi = \left[1 + \frac{\Lambda_{rec}(T)}{\Lambda_{ion}(T)}\right]^{-1} \approx \left[1 + 2.8 \times 10^{-6} T_5^{-7/6} e^{1.58/T_5}\right]^{-1}. \tag{6.13}$$

The observed absence of a Gunn-Peterson trough in the spectra of high-redshift quasars strongly constrains the density of neutral hydrogen in the IGM. The most thorough study to date, involving eight quasars (Steidel & Sargent 1987), concluded that

$$\Omega_{H_I}(z = 2.64) < (1.2 \pm 3.1) \times 10^{-8} h_{50}^{-1}$$

if $\Omega = 1$. In our model this corresponds to $(1 - \chi) < (1.2 \pm 3.1) \times 10^{-8} / (f_m \Omega_b) \approx 2 \times 10^{-6}$ for $\Omega_b = 0.06$ and f_m=0.1. Thus we are helped not only by the IGM being ionized, but also by it being diffuse. In a recent study of a single quasar, Webb *et al.* (1992) find the data consistent with either $\Omega_{H_I}(z = 4.1) = 0$ or $\Omega_{H_I}(z = 4.1) = 1.5 \times 10^{-8} h_{50}^{-1}$, depending on model assumptions. We will use the latter value as an upper limit. Finally, recent Hubble Space Telescope spectroscopy of 3C 273 has been used to infer that $\Omega_{H_I}(z = 0.158) < 1.4 \times 10^{-7} h_{50}^{-1}$. The constraints from these three studies are plotted in Figures 6.5 and 6.6 together with the ionization levels predicted by our scenario.

To achieve $\chi = 10^{-4}$, 10^{-5} and 10^{-6} in equilibrium would by equation (6.13) require $T > 5.5 \times 10^4\,K$, $T > 1.1 \times 10^5\,K$ and $T > 3.6 \times 10^5\,K$, respectively. As can be seen from the numerical solutions in the bottom panels of Figures 6.5 and 6.6, the recombination rate is generally too slow for equilibrium to be established, and the IGM remains almost completely ionized even when $T \ll 15,000K$ and equilibrium would have yielded $\chi \approx 0$. In both 3a and 3b, a very moderate reheating ($f_{inj} = 0.001$, heavy lines) is seen to suffice to satisfy the three observational constraints. In the absence of any reheating whatsoever, the only models that satisfy the constraints are those with very low density ($\Omega_b = 0.01$ or $f_m = 0.01$).

In summary, the only parameters that are strongly constrained by the Gunn-Peterson test are M_c and $\sigma_0(M_c)$.

6.3.5 Other spectral constraints

Let us estimate to what extent Compton cooling of the hot plasma will distort the cosmic microwave background radiation (CBR). Since for $T_e \gg T_\gamma$ the Comptonization y-parameter (Stebbins & Silk 1986)

$$y_C \equiv \int_t^{t_0} \frac{kT_e}{m_e c^2} n_e \sigma_t c\, dt$$

is linear in the plasma energy density $\left(\frac{3}{2} + \frac{3}{2}\right) kT_e n_e$ at each fixed time ($T_e = T_{IGM}$), all that counts is the spatially averaged thermal energy density at each redshift. Since the former is simply $\varepsilon_t(z;z_*)$ times the density of injected energy $f_g f_{sn}\Omega_b\rho c^2$, the calculation reduces to mere energetics and we obtain

$$y_C = y_* \int_z^\infty \frac{df_g(z_*)}{d(-z_*)} \int_0^{z_*} \sqrt{1+z}\,\varepsilon_t(z;z_*)dz\,dz_*,$$

where

$$y_* \equiv \frac{1}{8} f_{sn}\Omega_b^2 \frac{\sigma_t c H}{m_p G} \approx 9\times 10^{-7} h\Omega_b^2.$$

The current observational upper limit on y is 2.5×10^{-5} (Mather *et al.* 1994), so even if we take $f_g(0)$ as high as 100% and make a gross overestimate of the integral by making all our galaxies as early as at $z_* = 30$ and by replacing $\varepsilon_t(z;z_*)$ by its upper bound 60% for all z, z_*, our y is below the observational limit by three orders of magnitude for $\Omega = 1$.

Now let us estimate the optical depth of the IGM. It has long been known that reionization can cause a spatial smoothing of the microwave background as CBR photons Thomson scatter off of free electrons. Since $n_e = \chi_{IGM} n_b$, the optical depth for Thomson scattering, i.e. the number of mean free paths that a CBR photon has traversed when it reaches our detectors, is

$$\tau_t = \int_{t_{rec}}^{t_0} \sigma_t \chi_{IGM} n_b c\, dt = \tau_t^* \int_0^{z_{rec}} \sqrt{1+z}\chi_{IGM} dz,$$

where

$$\tau_t^* \equiv \frac{3}{8\pi} f_m \Omega_b \frac{H_0 c\sigma_t}{m_p G} \approx 0.07\Omega_b h.$$

Let us evaluate the integral by making the approximation that χ_{IGM} increases abruptly from 0 to 1 at some redshift z_{ion}. Then even for z_{ion} as high as 30, $\tau_t \approx 7.9h\Omega_b f_m \approx 0.02 \ll 1$ for our fiducial parameter values $h = 0.5, \Omega_b = 0.06$ and $f_m = 0.1$, so the probability that a given CBR photon is never scattered at all is $e^{-0.02} \approx 98\%$. Hence this scenario for late reionization will have only a very marginal smoothing effect on the CBR. If the shells are totally ionized as well, then the factor f_m disappears from the expressions above which helps only slightly. Then $z_{ion} = 15$ would imply that 8% of the CBR would be spatially smoothed on scales of a few degrees.

6.4 Discussion

We have calculated the effects of supernova driven winds from early galaxies assuming a Press-Schechter model of galaxy formation and a CDM power spectrum. The calculations have shown that reionization by such winds can indeed explain the observed absence of a Gunn-Peterson effect if a number of conditions are satisfied:

1. The masses of the first generation of galaxies must be very small, not greater than about $10^8 M_\odot$.

2. There is enough power on these small scales to get at least 10% of the baryons in galaxies by $z = 5$.

3. Except for the case where Ω_b is as low as 0.01, there must be some reheating of the IGM after $z = 5$ to prevent the IGM from recombining beyond allowed levels.

4. The commonly used thin-shell approximation for expanding bubbles must remain valid over cosmological timescales, with the mass fraction in the interior remaining much less than unity.

Whether 1) is satisfied or not depends crucially on the model for structure formation. This scenario is consistent with a pure CDM model and some low-bias tilted CDM models, but not with top-down models like pure HDM.

Observations of nearly solar abundances of heavy elements in intracluster gas have given some support for 2), which is roughly equivalent to requiring that at least 10% of the heavy elements in the universe be made before $z = 5$ (or whenever $\phi \gg 1$). As discussed in Section 6.3.2, the observations of some extremely metal-poor objects in QSO absorption line studies do not necessarily rule out our scenario, since it is highly uncertain whether all the hydrogen in the swept-up IGM would get thoroughly mixed with the metal-rich supernova ejecta. The fact that large numbers of mini-galaxies are not seen today need not be a problem either. Possible explanations for this range from mechanisms for physically destroying them (Dekel & Silk 1986, for instance) to the fact that the faint end of the luminosity function is still so poorly known that old dwarf galaxies in the field may be too faint to see by the present epoch (Binggeli *et al.* 1988).

To violate 3), the actual reheating would have to be extremely small. A current IGM temperature between $10^4 K$ and $10^5 K$ suffices, depending on other parameter values, since the low density IGM never has time to reach its equilibrium ionization.

The thin-shell approximation 4) is obviously a weak point in the analysis, because of the simplistic treatment of the dense shell and its interface with the interior bubble. For instance, could the shell cool and fragment due to gravitational instability before it collides with other shells? An approximate analytic model for such instability has been provided by Ostriker & Cowie (1981). Their criterion is that instability sets in when $\Xi > 1$, where

$$\Xi \equiv \frac{2G\rho_{shell}R^2}{\dot{R}v_s}$$

and the sound speed $v_s = \sqrt{5kT/3m_p}$. In terms of our dimensionless variables, this becomes

$$\Xi \approx 0.011 \times M_5^{1/5} T_5^{-1/2} \left(\frac{\Omega_b}{0.06}\right) \left(\frac{\delta}{0.1}\right)^{-1} \left(\frac{1+z}{1+z_*}\right)^3 \frac{r^2}{r'},$$

which indicates that with our standard parameter values, gravitational instability does not pose problems even with fairly low shell temperatures. The reason that the shell density is not limited to four times the ambient IGM density is that the jump condition is not adiabatic, due mainly to effective Compton cooling at the high redshifts under consideration.

After the critical z (typically between 20 and 5) at which the expanding shells have collided with neighbors and occupied most of space, the IGM is "frothy" on scales around 100 kpc, with dense cool shell remnants scattered in a hot thin and fairly uniform plasma. Since the dark matter distribution is left almost unaffected by the expanding bubbles, formation of larger structures such as the galaxies we observe today should remain fairly unaffected as far as concerns gravitational instability. There is indirect influence, however: the ubiquitous metals created by the early mini-galaxies would enhance the ability of the IGM to cool, which as mentioned in Section 6.3.1 is commonly believed to be crucial for galaxy formation.

Blanchard *et al.* (1992) argue that if the IGM has a temperature higher than the virial temperature of a dark halo, pressure support will prevent it from falling into the potential well and thus stop it from forming a luminous galaxy. The virial temperature they estimate for an object of mass M formed at a redshift z is approximately

$$T_{vir} \approx 5.7 \times 10^5 K \left(\frac{M}{10^{12} M_\odot}\right)^{2/3} (1+z)$$

for $h = 0.5$, so requiring $T_{vir} > T_{IGM}$ for say $T_{IGM} = 10^6 K$ at $z = 5$ would give a minimum galaxy mass of about $10^{11} M_\odot$. Such arguments indicate that the IGM reheating of our scenario might produce a "mass desert" between the earliest mini-galaxies and the galaxies we see today: The first generation of galaxies, mini-galaxies with masses of perhaps 10^6 or $10^8 M_\odot$, would keep forming until their expanding bubbles had occupied most of space and altered the bulk properties of the IGM. After that, formation of galaxies much smaller than than those of today would be suppressed, since the IGM would be too hot. Eventually, as the IGM cools by adiabatic expansion, a progressively larger fraction of the IGM can be accreted by dark matter potential wells. Indeed, even with the volume averaged IGM temperature remaining hot due to some form of reheating, cooling flows in deep potential wells, in particular galaxy clusters, would not be suppressed. Late formation of galaxies is therefore possible.

Pressure balance between the shell and the interior during the expansion would give the ration $T_{shell}/T_{interior} = \rho_{b,interior}/\rho_{b,shell} \approx 3\delta f_m \approx 0.03$ for $\delta = f_m = 0.1$, so the shell fragments would expected to contain non-negligible fractions of neutral hydrogen and thus absorb some Lyman-alpha. A typical shell radius is about 100 kpc for $M_c = 10^6 M_\odot$, a size comparable to that of the clouds of the Lyman-alpha forest. As to the number density of Lyman-alpha clouds, the observed velocity separations greatly exceed those we would expect if all shell fragments were to be identified with Lyman alpha clouds. Thus the majority of these fragments must have been destroyed by some other process. There are a number of ways in which this could occur, for instance through photoionization by UV flux from the parent galaxy or by collapse to form other dwarf galaxies. The resulting numbers resemble the abundance of minihalos in an alternative explanation of the Lyman-alpha forest (Rees 1986). Strong evolution, in the sense of an increasing cloud abundance with decreasing redshift, is expected to occur as cooling becomes effective.

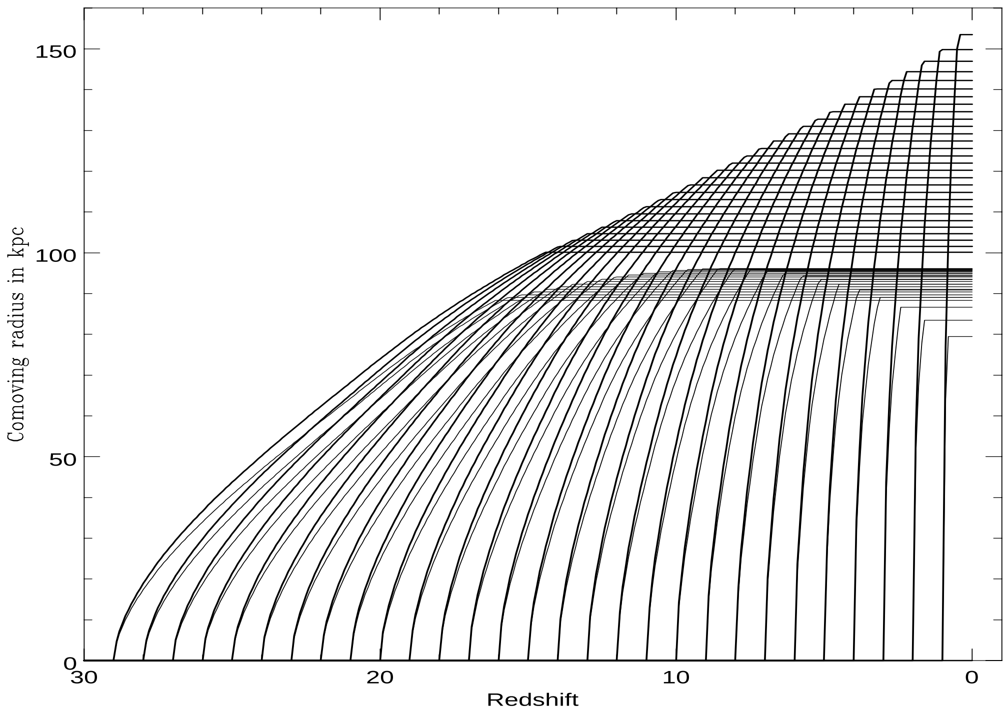


Figure 6.1: Comoving radius of expanding shell.

The comoving shell radius $(1+z)R$ is plotted for galaxies of total mass $2 \times 10^6 M_\odot$, forming at integer redshifts from 1 to 29. Here $\Omega = 1$, $\Omega_b = 0.06$, $h = 0.5$, and $f_m = 0.1$. $f_d = 1$ for the upper set of lines and $f_d = 0$ for the lower set. R has been truncated when T drops below 15,000 K, after which newly swept up IGM fails to become ionized.

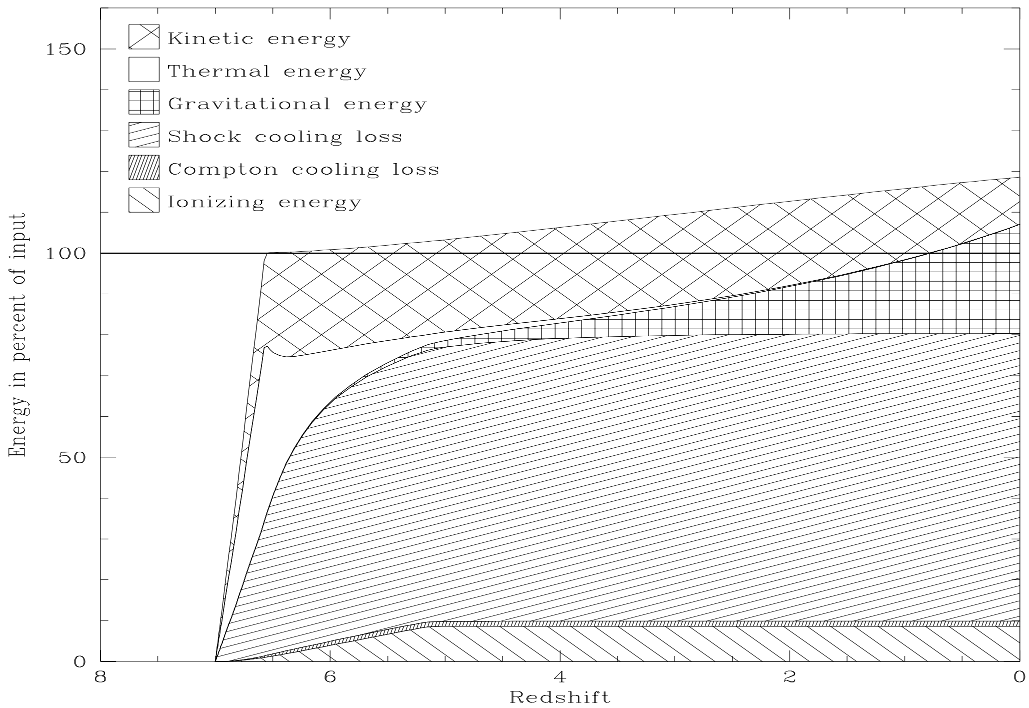


Figure 6.2: Energetics of expanding shell, example 1.

This and the two following figures show the energy contents of an expanding bubble as a function of redshift, for different choices of f_d and z_*. $\Omega = 1$, $\Omega_b = 0.06$, $h = 0.5$ and $f_m = 0.1$ for all three plots. Figures 6.2 and 6.3 illustrate the difference between $f_d = 0$ and $f_d = 1$ (there is no shock cooling loss in the second case). Figure 6.4 has $f_d = 0$ and illustrates that the Compton cooling loss is larger at higher redshift.

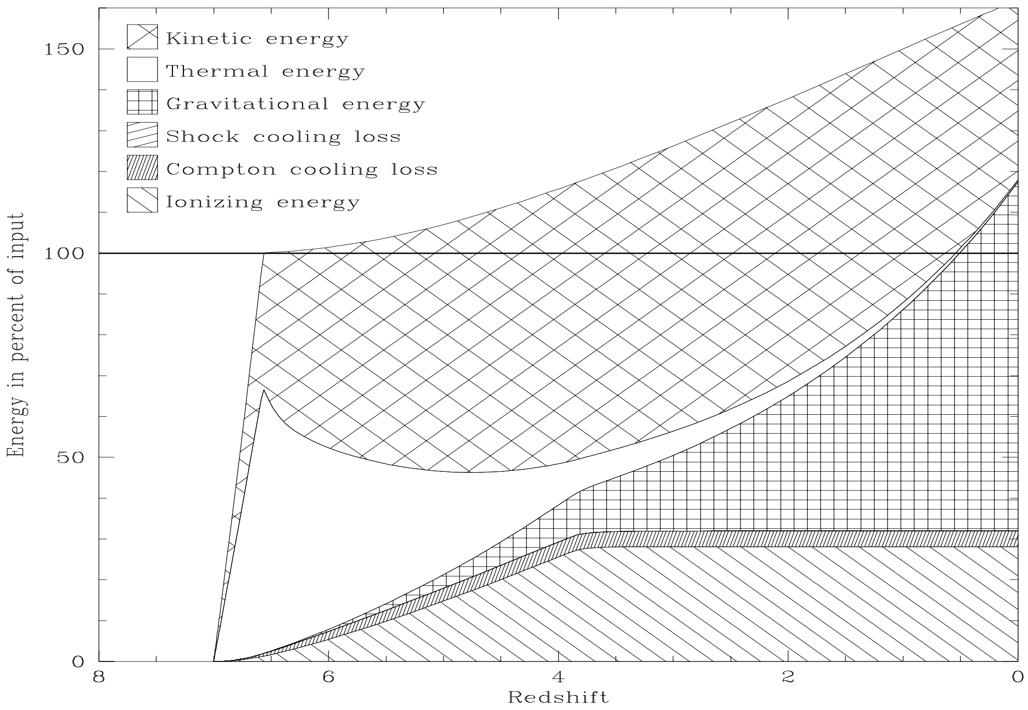


Figure 6.3: Energetics of expanding shell, example 2.

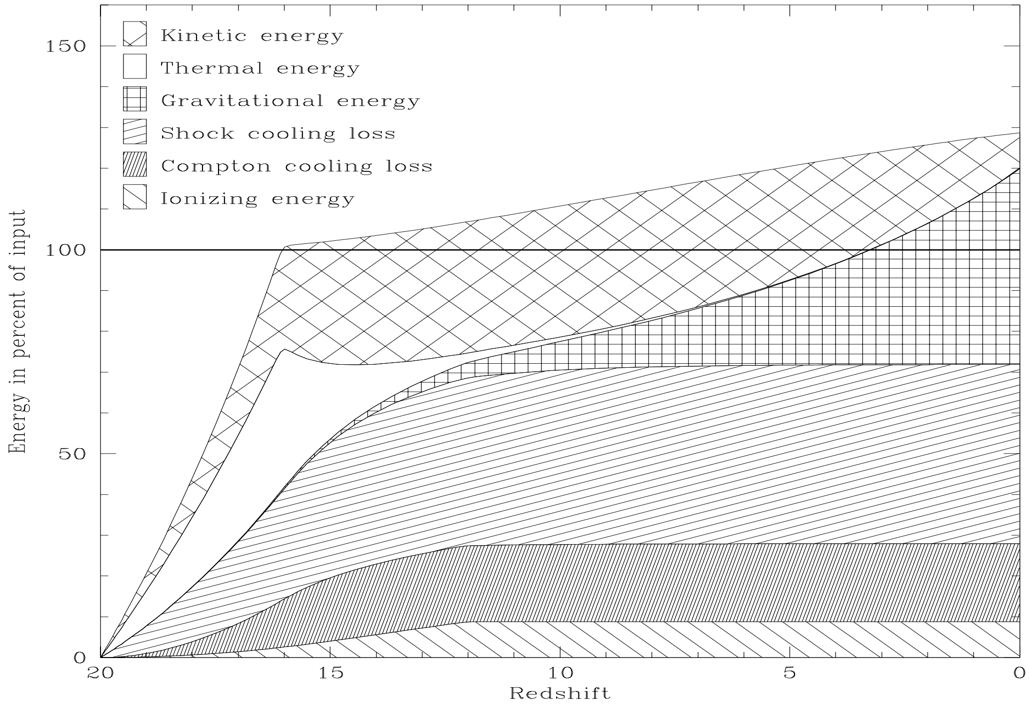
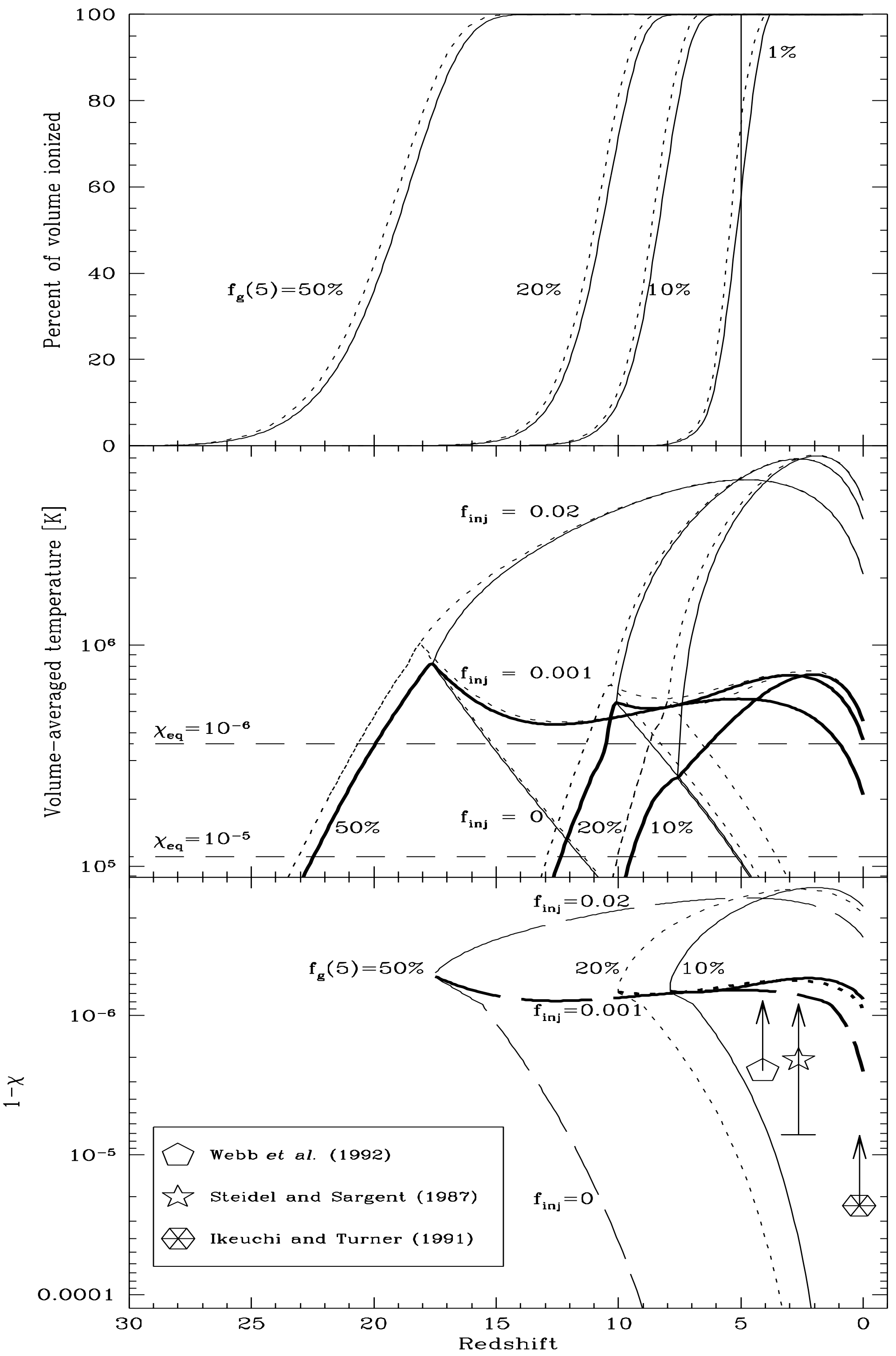
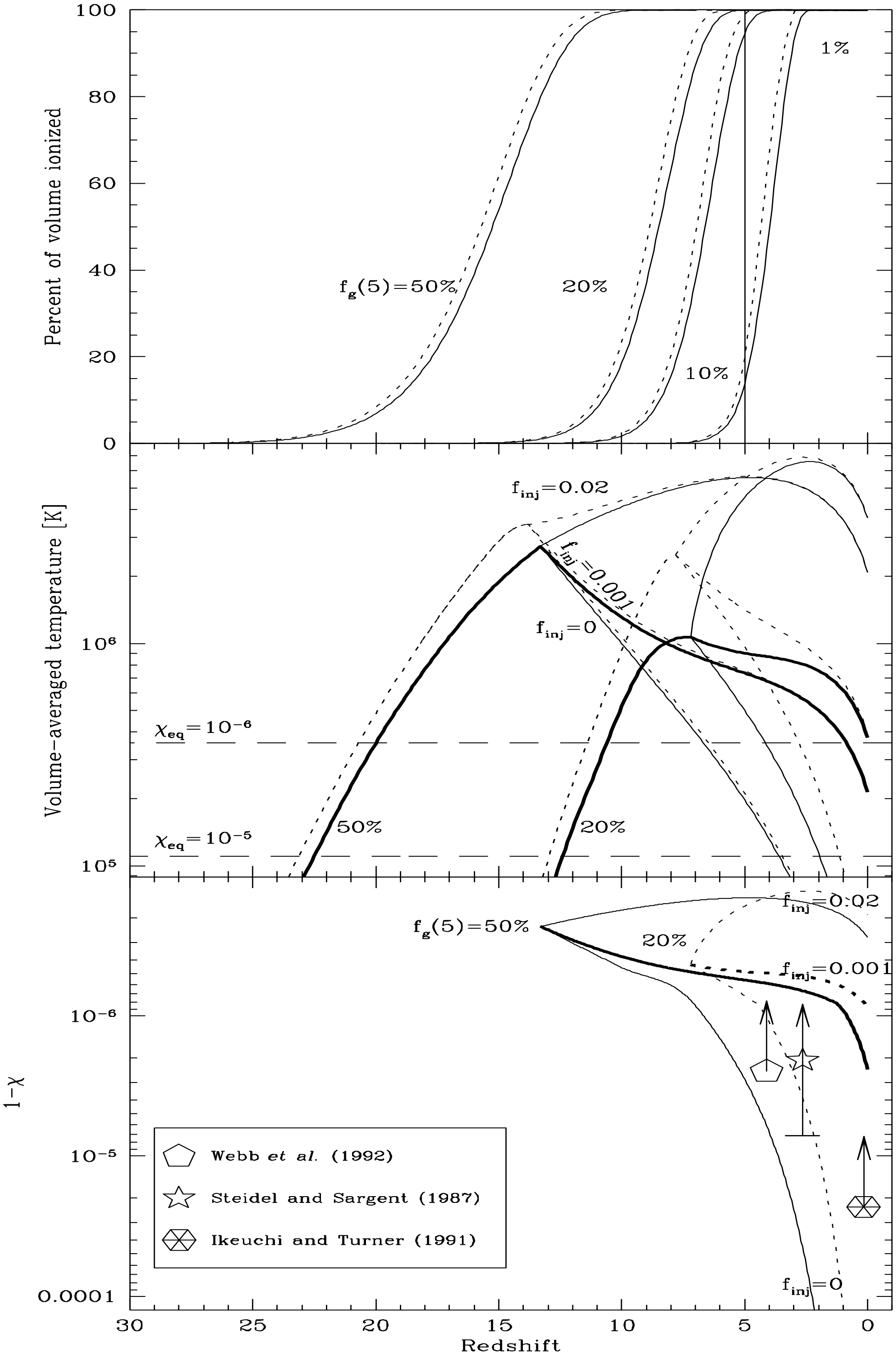


Figure 6.4: Energetics of expanding shell, example 3.



100
80
60
40
20
0
Percent of volume ionized
$f_g(5)=50\%$
20%
10%
1%
Volume-averaged temperature [K]
10^6
10^5
$f_{inj} = 0.02$
$f_{inj} = 0.001$
$f_{inj} = 0$
$\chi_{eq}=10^{-6}$
$\chi_{eq}=10^{-5}$
50%
20%
10%
$1-\chi$
10^{-6}
10^{-5}
0.0001
$f_{inj}=0.02$
$f_g(5)=50\%$
20%
10%
$f_{inj}=0.001$
$f_{inj}=0$
Webb et al. (1992)
Steidel and Sargent (1987)
Ikeuchi and Turner (1991)
30
25
20
15
10
5
0
Redshift


Percent of volume ionized
Volume-averaged temperature [K]
1-χ
Redshift
fg(5)=50%
20%
10%
1%
finj=0.02
finj=0.001
finj=0
χeq=10−6
χeq=10−5
50%
Webb et al. (1992)
Steidel and Sargent (1987)
Ikeuchi and Turner (1991)

Figures 6.5 and 6.6: IGM evolution.

Three different properties of the IGM (filling factor, temperature and ionization) are plotted as a function of redshift for different choices of M_c, $f_g(5)$ and f_d. $\Omega = 1$, $\Omega_b = 0.06$, $h = 0.5$ and $f_m = 0.1$ in the two previous figures, 6.5 and 6.6. In all panels, the different families of curves correspond to different values of $f_g(5)$; 50%, 20%, 10% and 1% from left to right, with the rightmost cases being omitted where they fail dismally. In the porosity and temperature plots (the upper two panels of 6.5 and 6.6), dashed lines correspond to $f_d = 1$ and solid ones to $f_d = 0$, whereas only the pessimistic $f_d = 0$ case is plotted in the ionization plots (the lower third). In the temperature and ionization plots (the lower two panels), the three branches of each curve correspond to the three reheating scenarios: $f_{inj} = 0$, $f_{inj} = 0.001$ and $f_{inj} = 0.02$.

Chapter 7

Power Spectrum Independent Constraints

In this chapter, a formalism is presented that allows cosmological experiments to be tested for consistency, and allows a simple frequentist interpretation of the resulting significance levels. As an example of an application, this formalism is used to place constraints on bulk flows of galaxies using the results of the microwave background anisotropy experiments COBE and SP91, and a few simplifying approximations about the experimental window functions. It is found that if taken at face value, with the quoted errors, the recent detection by Lauer and Postman of a bulk flow of 689 km/s on scales of $150h^{-1}$Mpc is inconsistent with SP91 at a 95% confidence level within the framework of a Cold Dark Matter (CDM) model. The same consistency test is also used to place constraints that are completely model-independent, in the sense that they hold for any power spectrum whatsoever — the only assumption being that the random fields are Gaussian. It is shown that the resulting infinite-dimensional optimization problem reduces to a set of coupled non-linear equations that can readily be solved numerically. Applying this technique to the above-mentioned example, we find that the Lauer and Postman result is inconsistent with SP91 even if no assumptions whatsoever are made about the power spectrum.

7.1 Introduction

Together with the classical cosmological parameters h, Ω, *etc.*, the power spectrum $P(k)$ of cosmological density fluctuations is one of the most sought-for quantities in modern cosmology, vital for understanding both the formation of large-scale structure and the fluctuations in the cosmic microwave background radiation (CBR).

The traditional approach has been to assume some functional form for $P(k)$ (like that predicted by the cold dark matter (CDM) scenario, for instance), and then investigate whether the predictions of the model are consistent with experimental data or not. The large amounts of data currently being produced by new CBR experiments and galaxy surveys, all probing different parts of the power spectrum, allow a new and more attractive approach. We can now begin to probe exact shape of the function $P(k)$, without making any prior assumptions about $P(k)$. More specifically, we measure different weighted averages of the function, the weights being the experimental window functions.

This new approach is quite timely (Juszkiewicz 1993), as there are now many indications that the primordial power spectrum may have been more complicated than an $n = 1$ power law. There are several sources of concern about the standard CDM cosmology, with inflation leading to $\Omega \approx 1$ and a primordial $n \approx 1$ Harrison-Zel'dovich power spectrum. Compared to COBE-normalized CDM, observational data shows unexpected large-scale bulk flows (Lauer & Postman 1994), too weak density correlations on small scales (Maddox *et al.* 1990), a rather quiet local velocity field (Schlegel *et al.* 1993) and a deficit of hot x-ray clusters (Oukbir & Blanchard 1992). The combined data from the COBE DMR (Smoot *et al.* 1992) and the Tenerife anisotropy experiment (Hancock *et al.* 1994) point to a spectral index exceeding unity (Watson & Gutiérrez de la Cruz 1993) which, if correct, cannot be explained by any of the standard inflationary models. The recent possible detections of halo gravitational microlensing events (Alcock *et al.* 1993) give increased credibility to the possibility that the dark matter in our galactic halo may be baryonic. If this is indeed the case, models with $\Omega < 1$ and nothing but baryonic dark matter (BDM) (Peebles 1987; Gnedin & Ostriker 1992, Cen, Ostriker & Peebles 1993) become rather appealing. However, in contrast to CDM with inflation, BDM models do not include a physical mechanism that makes a unique prediction for what the primeval power spectrum should be. Rather, the commonly assumed $P(k) \propto k^{-1/2}$ is chosen *ad hoc* to fit observational data. Moreover, for fluctuations near the curvature scale in open universes, where the $\Omega = 1$ Fourier modes are replaced by hyperspherical Bessel functions with the curvature radius as a built-in length scale, the whole notion of scale-invariance loses its meaning (Kamionkowski & Spergel 1993).

In summary, it may be advisable to avoid theoretical prejudice as to the shape of the primordial power spectrum. In this spirit, we will develop a consistency test that requires no such assumptions whatsoever about the form of the power spectrum. This approach was pioneered by Juszkiewicz; Górski and Silk (1987), who developed a formalism for comparing two experiments in a power-spectrum independent manner. We generalize this method to the case of more than two experiments, and then use the formalism to assess the consistency of three recent observational results: the CBR anisotropy measurements made by the COBE Differential Microwave Radiometer (Smoot *et al.* 1992), the South Pole anisotropy experiment (SP91, Gaier *et al.* 1992), and the measurement of bulk velocity of Abell clusters in a 150 h^{-1} Mpc sphere (Lauer & Postman 1994, hereafter LP).

In Section 7.2, we develop a formalism for testing cosmological models for consistency. In Section 7.3, we apply this formalism to the special case of cold dark matter (CDM) and the LP, SP91 and COBE experiments. In Section 7.4, we solve the variational problem that arises in consistency tests of models where we allow arbitrary power-spectra, and apply these results to the LP, SP91 and COBE experiments. Section 7.5 contains a discussion of our results. Finally, two different goodness-of-fit parameters are compared in Appendix A, and the relevant window functions are derived in Appendix B.

7.2 Consistency Tests for Cosmological Models

In cosmology, a field where error bars tend to be large, conclusions can depend crucially on the probabilistic interpretation of confidence limits. Confusion has sometimes arisen from the fact that large-scale measurements of microwave background anisotropies and bulk flows are fraught with two quite distinct sources of statistical uncertainty, usually termed experimental noise and cosmic variance. In this section, we present a detailed prescription for testing any model for consistency with experiments, and discuss the appropriate probabilistic interpretation of this test. By *model* we will mean not merely a model for the underlying physics, which predicts the physical quantities that we wish to measure, but also a model for the various experiments. Such a model is allowed to contain any number of free parameters. In subsequent sections, we give examples of both a very narrow class of models (standard CDM where the only free parameter is the overall normalization of the power spectrum), and a wider class of models (gravitational instability with Gaussian adiabatic fluctuations in a flat universe with the standard recombination history, the power spectrum being an arbitrary function).

Suppose that we are interested in N physical quantities $c_1, ..., c_N$, and have N experiments $E_1, ..., E_N$ devised such that the experiment E_i measures the quantity c_i. Let s_i denote the number actually obtained by the experiment E_i. Because of experimental noise, cosmic variance, *etc.*, we do not expect s_i to exactly equal c_i. Rather, s_i is a random variable that will yield different values each time the experiment is repeated. By repeating the experiment M times on this planet and averaging the results, the uncertainty due to experimental noise can be reduced by a factor $\sqrt{M}$. However, if the same experiment were carried out in a number of different horizon volumes throughout the universe (or, if we have ergodicity, in an ensemble of universes with different realizations of the underlying random field), the results would also be expected to differ. This second source of uncertainty is known as cosmic variance. We will treat both of these uncertainties together by simply requiring the model to specify the probability distribution for the random variables s_i.

Let us assume that the random variables s_i are all independent, so that the joint probability distribution is simply the product of the individual probability distributions, which we will denote $f_i(s)$. This is an excellent approximation for the microwave background and bulk flow experiments we will consider. Finally, let $\hat{s}_1, ..., \hat{s}_N$ denote the numbers actually obtained in one realization of the experiments.

The general procedure for statistical testing will be as follows:

- First, define a parameter η that is some sort of measure of how well the observed data s_i agree with the probability distributions f_i, with higher η corresponding to a better fit.

- Then compute the probability distribution $f_\eta(\eta)$ of this parameter, either analytically or by employing Monte-Carlo techniques.

- Compute the observed value of η, which we will denote $\hat{\eta}$.

- Finally, compute the probability $P(\eta < \hat{\eta})$, *i.e.* the probability of getting as bad agreement as we do or worse.

We will now discuss these four steps in more detail.

7.2.1 Choosing a goodness-of-fit parameter

Obviously, the ability to reject models at a high level of significance depends crucially on making a good choice of goodness-of-fit parameter η. In the literature, a common choice is the *likelihood product*, *i.e.*

$$\eta_l \propto \prod_{i=1}^{N} f_i(s_i).$$

In this chapter, we will instead use the *probability product*, *i.e.* the product of the probabilities P_i that each of the experiments yield results at least as extreme as observed. Thus if the observed $\hat{s}_i$ is smaller than the median of the distribution f_i, we have $P_i = 2P(s_i < \hat{s}_i)$, whereas $\hat{s}_i$ larger than the median would give $P_i = 2P(s_i > \hat{s}_i)$. The factor of two is present because we want a two-sided test. Thus $P_i = 1$ if $\hat{s}_i$ equals the median, $P_i = 2\%$ if $\hat{s}_i$ is at the high 99th percentile, *etc.*

These two goodness-of-fit parameters are compared in Appendix A for a few explicit examples, and the conclusions can be summarized as follows:

1. Rather unphysical probability distributions can be concocted that "fool" either one of these two parameters but not the other. Thus neither the likelihood product nor the probability product can be hailed as fundamentally better than the other.

2. For probability distributions encountered in the type of cosmology applications discussed in this chapter, always smooth and unimodal functions, the likelihood product and the probability product yield very similar results.

3. As described in the following sections, the probability distribution of the probability product can always be computed analytically. The probability distribution for the likelihood product, however, depends on the distributions of the underlying random variables, and except in a few fortuitous simple cases, it must be computed numerically through either repeated convolutions or Monte Carlo simulations. Thus the probability product is considerably more convenient to use.

Since 1) and 2) indicate that neither of the two goodness-of-fit parameters is preferable over the other on scientific grounds, the authors feel that 3) tips the balance in favor of the probability product.

7.2.2 Its probability distribution

Apart from the simple interpretation of the probability product η, it has the advantage that its probability distribution can be calculated analytically, and is completely independent of the physics of the model — in fact, it depends only on N. We will now give the exact distributions.

Confidence level	$N=1$	$N=2$	$N=3$	$N=4$
95%	0.05	0.0087	0.0018	0.00043
99%	0.01	0.0013	0.00022	0.000043
99.9%	0.001	0.000098	0.000013	0.0000021

Table 7.1: Probability product limits

By construction, $0 \leq \eta \leq 1$. For $N = 1$, η will simply have a uniform distribution, *i.e.*

$$f_{\mathbf{x}}(\eta) = \begin{cases} 1 & \text{if } 0 \leq \eta \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

Thus in the general case, η will be a product of N independent uniformly distributed random variables. The calculation of the probability distribution for η is straightforward, and can be found in a number of standard texts. The result is

$$f_{\mathbf{x}}(\eta) = -f_z(-\ln\eta)\frac{dz}{d\eta} = \begin{cases} \frac{1}{(N-1)!}(-\ln\eta)^{N-1} & \text{if } 0 \leq \eta \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

7.2.3 The consistency probability

The probability $P(\eta < \hat{\eta})$, the probability of getting as bad agreement as we do or worse, is simply the cumulative distribution function $F_\eta(\hat{\eta})$, and the integral can be carried out analytically for any N:

$$F_\eta(\hat{\eta}) \equiv P(\eta < \hat{\eta}) = \int_0^{\hat{\eta}} f_{\mathbf{x}}(u)du = \hat{\eta}\theta(\hat{\eta})\sum_{n=0}^{N-1}\frac{(-\ln\hat{\eta})^n}{n!}, \tag{7.1}$$

where θ is the Heaviside step function, and $F_\eta(\hat{\eta}) = 1$ for $\hat{\eta} \geq 1$. Since the product of N numbers between zero and one tends to zero as $N \to \infty$, it is no surprise that

$$F_\eta(\hat{\eta}) \to \theta(\hat{\eta})\hat{\eta}e^{-\ln\hat{\eta}} = \theta(\hat{\eta})$$

as $N \to \infty$, *i.e.* that $f_{\mathbf{x}}(\hat{\eta}) \to \delta(\hat{\eta})$. The function $F_\eta(\hat{\eta})$ is plotted in Figure 7.1, and the values of $\hat{\eta}$ for which $F_\eta(\hat{\eta}) = 0.05$, 0.01 and 0.001, respectively, are given in Table 7.1 for a few N-values. For example, if three experimental results give a goodness-of-fit parameter $\hat{\eta} = 0.0002$ for some model, then this model is ruled out at a confidence level of 99%. Thus if the model where true and the experiments where repeated in very many different horizon volumes of the universe, such a low goodness-of-fit value would be obtained less than 1% of the time.

7.2.4 Ruling out whole classes of models

If we wish to use the above formalism to test a whole set of models, then we need to solve an optimization problem to find the one model in the set for which the consistency probability is maximized. For instance, if the family of models under consideration is standard $n = 1$, $\Gamma = 0.5$ CDM (see Section 7.3), then the only free parameter is the overall normalization constant A. Thus we can write the consistency probability as $p(A)$, and use some numerical method to find the normalization A_* for which $p(A)$ is maximized. After this, the statistical interpretation is clear: if the experiments under consideration are carried out in an ensemble of CDM universes, as extreme results as those observed will only be obtained at most a fraction $p(A_*)$ of the time, whatever the true normalization constant is. Precisely this case will be treated in the next section. For the slightly wider class of models consisting of CDM power spectra with arbitrary A, n and Γ, the resulting optimization problem would be a three-dimensional one, and the maximal consistency probability would necessarily satisfy

$$p(A_*, n_*, \Gamma_*) \geq p(A_*, 1, 0.5) = p(A_*).$$

An even more general class of models is the set of all models where the random fields are Gaussian, *i.e.* allowing completely arbitrary power spectra P. In section 4, we will show that the resulting infinite-dimensional optimization problem can in be reduced to a succession of two finite-dimensional ones.

7.3 Cold Dark Matter Confronts SP91, COBE and Lauer-Postman

As an example of an application of the formalism presented in the previous section, we will now test the standard cold dark matter (CDM) model of structure formation for consistency with the SP91 CBR experiment and the Lauer-Postman bulk flow experiment.

Let E_1 be the Lauer-Postman (LP for short) measurement of bulk flows of galaxies in a $150h^{-1}$Mpc sphere (Lauer& Postman, 1994). Let E_2 be the 1991 South Pole CBR anisotropy experiment, SP91 for short (Gaier *et al.* 1992). Let E_3 be the COBE DMR experiment (Smoot *et al.* 1992). All of these experiments probe scales that are well described by linear perturbation theory, and so as long as the initial fluctuation are Gaussian, the expected results of the experiments can be expressed simply as integrals over the power spectrum of the matter perturbation:

$$\langle s_i \rangle = \int W_i(k) P(k) dk.$$

Here s_{sp} and s_{cobe} are the mean-square temperature fluctuations measured by the experiments, and $s_{lp} \equiv (v/c)^2$ is the squared bulk flow. The corresponding window functions W_{lp}, W_{sp} and W_{cobe} are derived in Appendix B, and plotted in Figure 7.2. These window functions assume that the initial perturbations were adiabatic, that $\Omega = 1$, and that recombination happened in the standard way, x.e. a last-scattering surface at $z \approx 1000$. The SP91 window function is to be interpreted as a lower limit to the true window function, as it includes contributions only from the Sachs-Wolfe effect, not from Doppler motions or intrinsic density fluctuations of the surface of last scattering.

Now let us turn to the probability distributions for the random variables s_{lp}, s_{sp} and s_c. The standard CDM model with power-law initial fluctuations $\propto k^n$ predicts a power spectrum that is well fitted by (Bond & Efstathiou 1984)

$$P(k) = \frac{Aq^n}{\left(1 + [aq + (bq)^{1.5} + (cq)^2]^{1.13}\right)^{2/1.13}},$$

n	Γ	LP	SP91	COBE	LP/SP91	LP/COBE
1	0.5	9.2×10^{-7}	1.6×10^{-8}	2.0×10^{-8}	56.7	45.1
0.7	0.5	1.7×10^{-6}	2.9×10^{-8}	5.1×10^{-8}	57.2	32.7
2	0.5	1.4×10^{-7}	2.6×10^{-9}	1.2×10^{-9}	53.9	112.2
1	0.1	1.4×10^{-6}	2.3×10^{-8}	4.3×10^{-8}	57.9	31.9
1	10	2.3×10^{-7}	4.1×10^{-9}	4.6×10^{-9}	55.9	49.6

Table 7.2: Expected r.m.s. signals for CDM power spectrum with $A = (1h^{-1}\mathrm{Mpc})^3$

where $a = 6.4$, $b = 3.0$, $c = 1.7$ and $q = (1h^{-1}\mathrm{Mpc})k/\Gamma$. For the simplest model, $\Gamma = h$, but certain additional complications such as a non-zero cosmological constant Λ and a non-zero fraction Ω_ν of hot dark matter can be fitted with reasonable accuracy by other values of Γ (Efstathiou, Bond & White 1992). Thus the model has three free parameters: n, Γ and the overall normalization A. Integrating the power spectrum against the three window functions yields the values of c_i given in Table 7.2. The two rightmost columns contain the quotients c_{lp}/c_{sp} and c_{lp}/c_c, respectively. As can be seen, the dependence on Γ is quite weak, and the quotient c_{lp}/c_{sp} is quite insensitive to the spectral index n as well. Let us for definiteness assume the canonical values $n = 1$ and $\Gamma = 0.5$ in what follows.

These values c_i would be the average values of the probability distributions for s_{sp} and s_{lp} if there where no experimental noise. We will now model the full probability distributions of the three experiments, including the contribution from experimental noise.

For a bulk flow experiment, the three components v_x, v_y and v_z of the velocity vector $\mathbf{v}$ are expected to be independent Gaussian random variables with zero mean, and

$$\langle|\mathbf{v}|^2\rangle = c_{lp}.$$

However, this is not quite the random variable s_{lp} that we measure, because of errors in distance estimation, *etc.* Denoting the difference between the observed and true bulk velocity vectors by ϵ, let us assume that the three components of ϵ are identically distributed and independent Gaussian random variables. This should be a good approximation, since even if the errors for individual galaxies are not, the errors in the average velocity ϵ will be approximately Gaussian by the Central Limit Theorem. Thus the velocity vector that we measure, $\mathbf{v}+\epsilon$, is also Gaussian, being the sum of two Gaussians. The variable that we actually measure is $s_{lp} = |\mathbf{v}+\epsilon|^2$, so

$$s_{lp} = \frac{1}{3}\left(c_{lp} + V_{lp}\right)\chi_3^2,$$

where χ_3^2 has a chi-squared distribution with three degrees of freedom, and V_{lp} is the variance due to experimental noise, *i.e.* the average variance that would be detected even if the true power spectrum were $P(k) = 0$. The fact that the expectation value of the detected signal s_{lp} (which is usually referred to as the *uncorrected* signal in the literature) exceeds the true signal c_{lp} is usually referred to as *error bias* (LP; Strauss, Cen & Ostriker 1993 — hereafter SCO). Error bias is ubiquitous to all experiments of the type discussed in this chapter, including CBR experiments, since the measured quantity is positive definite and the noise errors contribute squared. In the literature, experimentally detected signals are usually quoted after error bias has been corrected for, *i.e.* after the noise has been subtracted from the uncorrected signal in For LP, the uncorrected signal is 807 km/s, whereas the signal quoted after error bias correction is 689 km/s.

For the special case of the LP experiment, detailed probability distributions have been computed using Monte-Carlo simulations (LP, SCO), which incorporate such experiment-specific complications as sampling errors, asymmetry in the error ellipsoid, *etc.* To be used here, such simulations would need to be carried out for each value of c_{lp} under consideration. Since the purpose of this section is merely to give an example of the test formalism, the above-mentioned χ^2-approximation will be quite sufficient for our needs.

For the SP91 nine-point scan, the nine true values $\Delta T_i/T$ are expected to be Gaussian random variables that to a good approximation are independent. They have zero mean, and

$$\langle|\Delta T_i/T|^2\rangle = c_{sp}.$$

Denoting the difference between the actual and observed values by δ_i, we make the standard assumption that these nine quantities are identically distributed and independent Gaussian random variables. Thus the temperature fluctuation that we measure at each point, $\Delta T_i/T+\delta$, is again Gaussian, being the sum of two Gaussians. The variable that we actually measure is

$$s_{sp} = \frac{1}{9}\sum_1^9\left(\frac{\Delta T_i}{T}+\delta_i\right)^2 = \frac{1}{9}\left(c_{sp}+V_{sp}\right)\chi_9^2,$$

where χ_9^2 has a chi-squared distribution with nine degrees of freedom, and V_{sp} is the variance due to experimental noise, the error bias, *i.e.* the average variance that would be detected even if the true power spectrum were $P(k)=0$.

We will use only the signal from highest of the four frequency channels, which is the one likely to be the least affected by galactic contamination. Again, although Monte-Carlo simulations would be needed to obtain the exact probability distributions, we will use the simple χ^2-approximation here. In this case, the main experiment-specific complication is the reported gradient removal, which is a non-linear operation and thus does not simply lead to a χ^2-distribution with fewer degrees of freedom.

The amplitude of the COBE signal can be characterized by the variance in $\Delta T/T$ on an angular scale of 10°. This number can be estimated from the COBE data set as $s_c = \sigma^2_{10^\circ} = ((11.0\pm1.8)\times10^{-5})^2$ (Smoot *et al.* 1992). The uncertainty in this quantity is purely due to instrument noise, and contains no allowance for cosmic variance. We must fold in the contribution due to cosmic variance in order to determine the probability distribution for s_c. We determined this probability distribution by performing Monte-Carlo simulations of the COBE experiment. We made simulated COBE maps with a variety of power spectra (including power laws with indices ranging from 0 to 3, as well as delta-function power spectra of the sort described in Section 7.2). We included instrumental noise in the maps, and excluded all points within 20° of the Galactic plane. By estimating s_c from each map, we were able to construct a probability distribution corresponding to each power spectrum.

In all cases, the first three moments of the distribution were well approximated by

$$\begin{cases} \mu_1 & \equiv \langle s_c \rangle = c_c, \\ \mu_2 & \equiv \langle s_c^2 \rangle - \langle s_c \rangle^2 \leq 0.063 c_c^2 + 1.44 \times 10^{-21}, \\ \mu_3 & \equiv \langle s_c^3 \rangle = 0.009 c_c^3. \end{cases} \quad (7.2)$$

Furthermore, in all cases the probability distributions were well modeled by chi-squared distributions with the number of degrees of freedom, mean, and offset chosen to reproduce these three moments. Note that the magnitude of the cosmic variance depends on the shape of the power spectrum as well as its amplitude. The inequality in the above expression for μ_2 represents the largest cosmic variance of any of the power spectra we tested. Since we wish to set conservative limits on models, we will henceforth assume that the cosmic variance is given by this worst-case value. Thus we are assuming that the random variable $(s_c - s_0)/\Delta s$ has a chi-squared distribution with δ degrees of freedom, where

$$\begin{cases} s_0 & = \mu_1 - 2\mu_2^2/\mu_3, \\ \Delta s & = \mu_3/4\mu_2, \\ \delta & = 8\mu_2^3/\mu_3^2. \end{cases} \quad (7.3)$$

The results obtained using these three probability distributions are summarized in Tables 7.3, 7.4 and 7.5. In 7.3 and 7.4, $N = 2$, and the question asked is whether LP is consistent with COBE and SP91, respectively. In Table 7.5, $N = 3$, and we test all three experiments for consistency simultaneously. In each case, the optimum normalization (proportional to the entries labeled "Signal") is different, chosen such that the consistency probability for the experiments under consideration is maximized. As can be seen, the last two tests rule out CDM at a significance level of 95%, *i.e.* predict that in an ensemble of universes, results as extreme as those we observe would be obtained less than 5% of the time. Note that using both COBE and SP91 to constrain LP yields a rejection that is no stronger than that obtained when ignoring COBE. In the latter case, the best fit is indeed that with no cosmological power at all, which agrees well with the observation of SCO that sampling variance would lead LP to detect a sizable bulk flow (before correcting for error bias) even if there where none.

7.4 Allowing Arbitrary Power Spectra

In this section, we will derive the mathematical formalism for testing results from multiple experiments for consistency, without making any assumptions whatsoever about the power spectrum. This approach was pioneered by Juszkiewicz *et al.* (1987) for the case $N = 2$. Here we generalize the results to the case of arbitrary N. Despite the fact that the original optimization problem is infinite-dimensional, the necessary calculations will be seen to be of a numerically straightforward type, the case of N independent constraints leading to nothing more involved than numerically solving a system of N coupled non-linear equations. After showing this, we will discuss some inequalities that provide both a good approximation of the exact results and a useful qualitative understanding of them.

	LP	COBE	Combined
Noise	420 km/s	9.8 μK	
Signal	169 km/s	33.8 μK	
Noise+Signal	453 km/s	35.2 μK	
Detected	807 km/s	35.2 μK	
$\hat{\eta}$	0.046	1.00	0.046
$P(\eta < \hat{\eta})$	0.046	1.00	0.19

Table 7.3: Are LP and COBE consistent with CDM?

	LP	SP91	Combined
Noise	420 km/s	26.4 μK	
Signal	0 km/s	0 μK	
Noise+Signal	420 km/s	26.4 μK	
Detected	807 km/s	19.9 μK	
$\hat{\eta}$	0.023	0.35	0.0079
$P(\eta < \hat{\eta})$	0.023	0.35	0.046

Table 7.4: Are LP and SP91 consistent with CDM?

	LP	SP91	COBE	Combined
Noise	420 km/s	26.4 μK	9.8 μK	
Signal	168 km/s	26.9 μK	33.8 μK	
Noise+Signal	452 km/s	37.7 μK	35.1 μK	
Detected	807 km/s	19.9 μK	35.2 μK	
$\hat{\eta}$	0.046	0.039	0.97	0.0017
$P(\eta < \hat{\eta})$	0.046	0.039	0.97	0.046

Table 7.5: Are LP, SP91 and COBE all consistent with CDM?

Tables 7.3, 7.4 and 7.5 show the consistency probability calculations. The first line in each table gives the experimental noise, *i.e.* the detection that would be expected in the absence of any cosmological signal. The second line is the best-fit value for the cosmological signal c, the value that maximizes the combined consistency probability in the lower right corner of the table. The third line contains the expected value of an experimental detection, and is the sum in quadrature of the two preceding lines. The fourth line gives the goodness-of-fit parameter for each of the experiments, i.e. the probability that they would yield results at least as extreme as they did. The rightmost number is the combined goodness-of-fit parameter, which is the product of the others. The last line contains the consistency probabilities, the probabilities of obtaining goodness-of-fit parameters at least as low as those on the preceding line.

7.4.1 The optimization problem

Let us consider $N = n + 1$ experiments numbered 0, 1, ..., n that probe the cosmological power spectrum $P(k)$. We will think of each experiment as measuring some weighted average of the power spectrum, and characterize an experiment E_i by its window function $W_i(k)$ as before.

Purely hypothetically, suppose we that we had repeated the same experiments in many different locations in the universe, and for all practical purposes knew the quantities $c_1, ..., c_n$ exactly. Then for which power spectrum $P(k)$ would c_0 be maximized, and what would this maximum be? If we experimentally determined c_0 to be larger than this maximum value, our results would be inconsistent, and we would be forced to conclude that something was fundamentally wrong either with our theory or with one of the experiments. In this section, we will solve this hypothetical problem. After this, it will be seen that the real problem, including cosmic variance and experimental noise, can be solved in almost exactly the same way.

The extremal power spectrum we are looking for is the solution to the following linear variational problem:

Maximize

$$\int_0^\infty P(k)W_0(k)dk \tag{7.4}$$

subject to the constraints that

$$\begin{cases} \int_0^\infty P(k)W_i(k)dk = c_i & \text{for } = 1, ..., n, \\ P(k) \geq 0 & \text{for all } k \geq 0. \end{cases}$$

This is the infinite-dimensional analogue of the so called linear programming problem, and its solution is quite analogous to the finite-dimensional case. In geometrical terms, we think of each power spectrum as a point in the infinite dimensional vector space of power spectra (tempered distributions on the positive real line, to be precise), and limit ourselves to the subset Ω of points where all the above constraints are satisfied. We have a linear function on this space, and we seek the point within the subset Ω where this function is maximized. We know that a differentiable functional on a bounded region takes its maximum either at an interior point, at which its gradient will vanish, or at a boundary point. In linear optimization problems like the one above, the gradient (here the variation with respect to P, which is simply the function W_0) is simply a constant, and will never vanish. Thus any maximum will always be attained at a boundary point. Moreover, from the theory of linear programming, we know that if there are n linear constraint equations, then the optimum point will be a point where all but at most n of the coordinates are zero. It is straightforward to generalize this result to our infinite-dimensional case, where each fixed k specifies a "coordinate" $P(k)$, and the result is that the solution to the variational problem is of the form

$$P(k) = \sum_{i=1}^{n} p_i \delta(k - k_i).$$

This reduces the optimization problem from an infinite-dimensional one to a $2n$-dimensional one, where only the constants p_i and k_i remain to be determined:

Maximize

$$\sum_{j=1}^{n} p_j W_0(k_j) \tag{7.5}$$

subject to the constraints that

$$\begin{cases} \sum_{j=1}^{n} p_j W_i(k_j) = c_i & \text{for } i = 1, ..., n, \\ p_i \geq 0 & \text{for } i = 1, ..., n. \end{cases}$$

This problem is readily solved using the method of Lagrange multipliers: defining the Lagrangian

$$L = \sum_{j=1}^{n} p_j W_0(k_j) - \sum_{i=1}^{n} \lambda_i \left[\sum_{j=1}^{n} p_j W_i(k_j) - c_i\right]$$

and requiring that all derivatives vanish leaves the following set of $3n$ equations to determine the $3n$ unknowns p_i, k_i and λ_i:

$$\begin{cases} W_0(k_i) - \sum_{j=1}^{n} \lambda_j W_j(k_i) & = 0, \\ \left[W_0'(k_i) - \sum_{j=1}^{n} \lambda_j W_j'(k_i)\right] p_i & = 0, \\ c_i - \sum_{j=1}^{n} p_j W_i(k_j) & = 0. \end{cases}$$

Introducing matrix notation by defining the k_i-dependent quantities $A_{ij} \equiv W_j(k_i)$, $B_{ij} \equiv W_j'(k_i)$, $a_i \equiv W_0(k_i)$ and $b_i \equiv W_0'(k_i)$ brings out the structure of these equations more clearly: If $p_i \neq 0$, then

$$\begin{cases} A\lambda & = \mathbf{a}, \\ B\lambda & = \mathbf{b}, \\ A^T\mathbf{p} & = \mathbf{c}. \end{cases}$$

If A and B are invertible, then eliminating λ from the first two equations yields the following system of n equations to be solved for the n unknowns k_1, ..., k_n:

$$A^{-1}\mathbf{a} = B^{-1}\mathbf{b}. \tag{7.6}$$

Although this system is typically coupled and non-linear and out of reach of analytical solutions for realistic window functions, solving it numerically is quite straightforward. A useful feature is that once this system is solved, $\mathbf{a}$, $\mathbf{b}$, A and B are mere constants, and the other unknowns are simply given by matrix inversion:

$$\begin{cases} \lambda & = A^{-1}\mathbf{a} \\ \mathbf{p} & = \left(A^{-1}\right)^T \mathbf{c} \end{cases}$$

Since the non-linear system (7.6) may have more than one solution, all solutions should be substituted back into (7.5) to determine which one is the global maximum. Furthermore, to make statements about the solution to our original optimization problem (7.4), we need to consider also the case where one or more of the n variables p_1,...,p_n vanish. If exactly m of them are non-vanishing, then without loss of generality, we may assume that these are the first m of the n variables. Thus we need to solve the maximization problem (7.5) separately for the cases where $P(k)$ is composed of n delta functions, $n-1$ delta functions, *etc.*, all the way down to the case where $P(k)$ is single delta function. These solutions should then be substituted back into (7.4) to determine which is the global maximum sought in our original problem. Thus the solutions depend on the window functions W_i and the signals c_i in the following way:

- From the window functions alone, we can determine a discrete and usually finite number of candidate wavenumbers k where delta functions can be placed.
- The actual signals c_i enter only in determining the coefficients of the delta functions in the sum, i.e. in determining what amount of power should be hidden at the various candidate wavenumbers.

If we have found an the optimal solution, then a small change in the signal vector $\mathbf{c}$ will typically result in a small change in $\mathbf{p}$ and no change at all in the number of delta functions in $P(k)$ or their location. If $\mathbf{c}$ is changed by a large enough amount, the delta functions may suddenly jump and/or change in number as a different solution of (7.5) takes over as global optimum or one of the coefficients p_i becomes negative, the latter causing the local optimum to be rejected for constraint violation. Thus within certain limits, we get the extremely simple result that for the optimal power spectrum $P(k)$,

$$c_0 = \int_0^\infty P(k)W_0(k)dk = \left(A^{-1}\mathbf{a}\right)\cdot\mathbf{c}.$$

Thus within these limits, c_0 depends linearly on the observed signal strengths c_i. This is exactly analogous to what happens in linear programming problems.

7.4.2 A useful inequality

Before proceeding further, we will attempt to provide a more intuitive understanding of the results of the previous section, and show how to determine how complicated a calculation is justified. For the special case of only a single constraint, *i.e.* $n = 1$, we obtain simply $P(k) = p_1\delta(k - k_1)$, where k_1 is given by

$$W_0'(k_1)W_1(k_1) = W_1'(k_1)W_0(k_1).$$

For the case of n constraints, let us define the functions

$$f_i \equiv \frac{W_0(k)}{W_1(k)}c_i.$$

Then we see that for $n = 1$, k_1 is simply the wavenumber for which the function f_1 is maximized, and that the maximum signal possible is simply $c_0 = f_1(k_1)$. Thus the maximum signal in experiment 0 that is consistent with the constraint from experiment i is obtained when the power is concentrated where the function f_i is large. In other words, if we want to explain a high signal c_0 in the face of low signals in several constraining experiments, then the best place to hide the necessary power from the i^{th} experiment is where f_i takes its maximum. These functions are plotted in Figure 7.4 for the experiments discussed in Section 7.3, the optimization problem being the search for the maximum LP signal that is consistent with the constraints from SP91 and COBE. For illustrative purposes, we here assume that c_{sp} and c_c are known exactly, and given by the detected signals $\hat{s}_{sp}^{1/2} \approx 19.9\,\mu$K and $\hat{s}_c^{1/2} \approx 33.8\,\mu$K (we will give a proper treatment of cosmic variance and noise in the following section). Using the $n = 1$ constraint for each constraining experiment separately, the smallest of the functions thus sets an upper limit to the allowed signal $c_0 = c_{sp}$. Thus the limit is given by the highest point in the hatched region in Figure 7.4, *i.e.*

$$c_0 \leq c_{max}^{(1)} \equiv \sup_k \min_i f_i(k).$$

We see that using the SP91 constraint alone, the LP signal would be maximized if all power were at $k \approx (940\text{Mpc})^{-1}$. Since this flagrantly violates the COBE constraint, the best place to hide the power is instead at $k \approx (100\text{Mpc})^{-1}$.

By using the above formalism to impose all the constraints at once, the allowed signal obviously becomes lower. If the constraints are equalities rather than inequalities, then this stronger limit can never lie below the value at $(k_* \approx 250\text{Mpc})^{-1}$, where $f_{sp}(k_*) = f_c(k_*)$, since this is the signal that would result from a power spectrum of the form $P(k) \propto \delta(k - k_*)$. Thus for the particular window functions in our example, where the constraint from the $n = 2$ calculation cannot be more than a factor $f_{sp}(80\text{Mpc})/f_{sp}(250\text{Mpc}) \approx 1.05$ stronger than the simple $n = 1$ limits, the latter are so close to the true optimum that they are quite sufficient for our purposes. If the constraints are upper limits rather than equalities, then the limit on c_0 is more relaxed, and is always the uppermost point in the hatched region, *i.e.* $c_{max}^{(1)}$.

7.4.3 Including noise and cosmic variance

To correctly handle cosmic variance and instrumental noise, we need to use the formalism developed in Section 7.2. Thus given the probability distributions for the various experimental results s_i, we wish to find the power spectrum for which the consistency probability η is maximized. This optimization problem, in which all experiments are treated on an equal footing, will be seen to lead directly to the asymmetric case above where the signal in one is maximized given constraints from the others. For definiteness, we will continue using the example with the LP, SP91 and COBE experiments. As seen in Section 7.2, the source of the low consistency probabilities is that $\hat{s}_{lp}$ is quite high when compared to $\hat{s}_{sp}$ and $\hat{s}_c$. Thus it is fairly obvious that for the power spectrum that maximizes the consistency probability, we will have $\hat{s}_{lp} > \langle s_{lp} \rangle$, whereas $\hat{s}_{sp} < \langle s_{sp} \rangle$ and $\hat{s}_c < \langle s_c \rangle$, so we can neglect power spectra that do not have this property. Let us first restrict ourselves to the subset of these power spectra for which $c_{lp} = D$ and $c_c = E$, where D and E are some constants. Then these power spectra all predict the same probability distributions for s_{lp} and s_c. The consistency probability η is clearly maximized by the power spectrum that maximizes $\langle s_{sp} \rangle$, and this will be a linear combination of one or two delta functions as shown in Section 7.4.1. The key point is that since the locations of these delta functions are independent of D and E (within the range discussed in 4.1), the infinite-dimensional optimization problem reduces to the following two simple steps:

1. Solve for the optimal number of delta functions m and their locations k_i as described in Section 7.4.1

2. Find the m coefficients p_i for which the power spectrum $P(k) = \sum_{i=1}^{m} p_i\delta(k - k_i)$ maximizes the consistency probability.

7.4.4 Power spectrum independent constraints on LP, SP91 and COBE

When applying the above consistency test to the LP, SP91 and COBE experiments, we obtain exactly the same consistency probability as in Table 7.4. The reason for this is that the optimal normalization turns out to be zero. This will obviously change if the LP error bars become smaller in the future. Thus dropping the CDM assumption does not improve the situation at all, which indicates that main source of the inconsistency must be something other than the CDM model.

In anticipation of future developments, consistency probabilities were also computed for a number of cases with less noise in the LP experiment. Comparing only LP and SP91, the optimum power spectrum has a delta function at $k \approx (941\mathrm{Mpc})^{-1}$. When including all three experiments, treating the COBE and SP91 constraints as upper limits, the optimum power spectrum has a single delta function at $k \approx (79\mathrm{Mpc})^{-1}$, so the addition of COBE strengthens the constraint only slightly, due to the flatness of f_{sp} in Figure 7.4. Interestingly, for all these cases with smaller LP error bars, consistency probabilities were found to be almost as low when allowing arbitrary power spectra as for the CDM case. This is again attributable to the flatness of f_{sp}, since weighted averages of a flat function are fairly independent of the shape of the weight function (here the power spectrum).

7.5 Discussion

We have developed a formalism for testing multiple cosmological experiments for consistency. As an example of an application, we have used it to place constraints on bulk flows of galaxies using the COBE and SP91 measurements of fluctuations in the cosmic microwave background. It was found that taken at face value, the recent detection by Lauer and Postman of a bulk flow of 689 km/s on scales of $150h^{-1}$Mpc is inconsistent with SP91 within the framework of a CDM model, at a significance level of about 95%. However, interestingly, this cannot be due solely to the CDM assumption, since the LP result was shown to be inconsistent with COBE and SP91 at the same significance level even when no assumptions whatsoever were made about the power spectrum. This leaves four possibilities:

1. The window functions are not accurate.

2. Something is wrong with the quoted signals or error bars for at least one of the experiments,

3. The observed fluctuations cannot be explained within the framework of gravitational instability and the Sachs-Wolfe effect.

4. The random fields are not Gaussian,

Case (1) could be attributed to a number of effects: If $\Omega \neq 1$, then both the calculation of the Sachs-Wolfe effect (which determines W_{sp} and W_{cobe}) and the growth of velocity perturbations (which determines W_{lp}) are altered. If the universe became reionized early enough to rescatter a significant fraction of all CBR photons, then small scale CBR anisotropies were suppressed, which would lower W_{sp}. A quantitative treatment of these two cases will be postponed to future work. Other possible causes of (1) include a significant fraction of the density perturbations being isocurvature (entropy) perturbations or tensor-mode perturbations (gravity waves). Apart from these uncertainties, we have made several simplifying assumptions about the window functions for LP and SP91. To obtain more accurate consistency probabilities than those derived in this chapter, a more accurate LP window function should be used that incorporates the discreteness and the asymmetry of the sample of Abell clusters used. This can either be done analytically (Feldman & Watkins 1993) or circumvented altogether by performing Monte-Carlo simulations like those of LP or SCO, but for the whole family of power spectra under consideration.

As to case (2), there has been considerable debate about both the LP and the SP91 experiments. A recent Monte-Carlo Simulation of LP by SCO basically confirms the large error bars quoted by LP. As is evident from the flatness of the LP curve in Figure 7.3, it will be impossible to make very strong statements about inconsistency until future experiments produce smaller error bars. With the SP91 experiment, a source of concern is the validity of using only the highest of the four frequency channels to place limits, even though it is fairly clear that the other three channels suffer from problems with galactic contamination. The situation is made more disturbing by the fact that a measurement by the balloon-borne MAX experiment (Gundersen *et al.* 1993) has produced detections of degree-scale fluctuations that that are higher than those seen by SP91, and also higher than another MAX measurement (Meinhold *et al.* 1993). Other recent experiments that have detected greater fluctuations include ARGO (de Bernardis *et al.* 1993), PYTHON (Dragovan *et al.* 1993) and MSAM (Cheng *et al.* 1993). On the other hand, SP91 has been used only as an upper limit in our treatment, by including only the Sachs-Wolfe effect and neglecting both Doppler contributions from peculiar motions of the surface of last scattering and intrinsic density fluctuations at the recombination epoch. If these effects (which unfortunately depend strongly on parameters such as h and Ω_b) where included, the resulting constraints would be stronger.

Case (3) might be expected if the universe underwent a late-time phase transition, since this could generate new large-scale fluctuations in an entirely non-gravitational manner.

In the light of the many caveats in categories (1) and (2), the apparent inconsistency between LP and SP91 (Jaffe *et al.* 1993) is hardly a source of major concern at the present time, and it does not appear necessary to invoke (3) or (4). However, we expect the testing formalism developed in this chapter to be able to provide many useful constraints in the future, as more experimental data is accumulated and error bars become smaller.

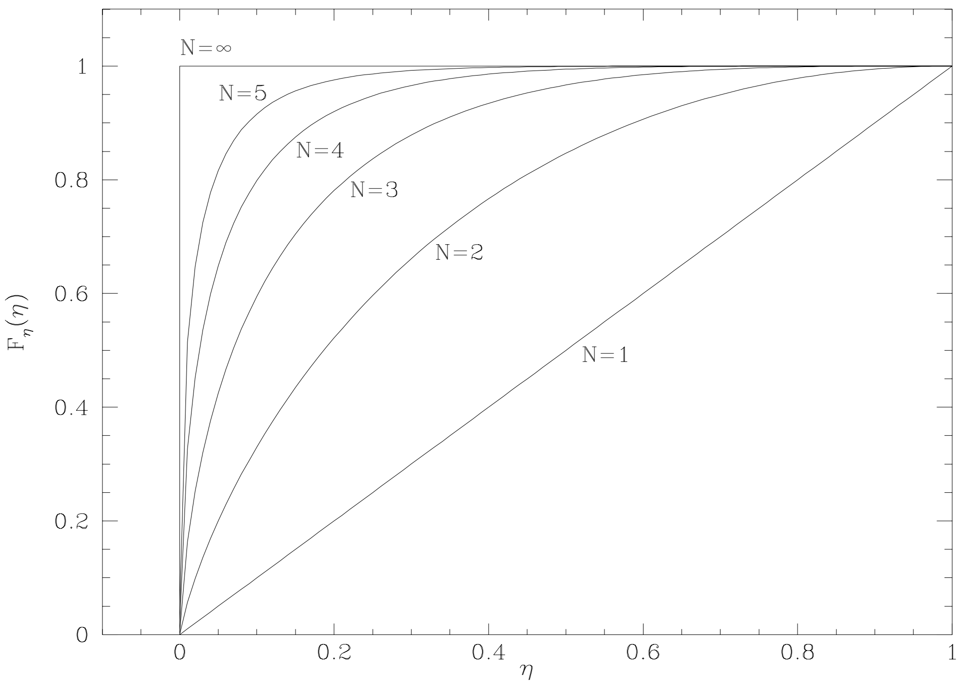


Figure 7.1: The function F_η.

The cumulative probability distribution for the goodness-of-fit parameter η is plotted for a few different n-values.

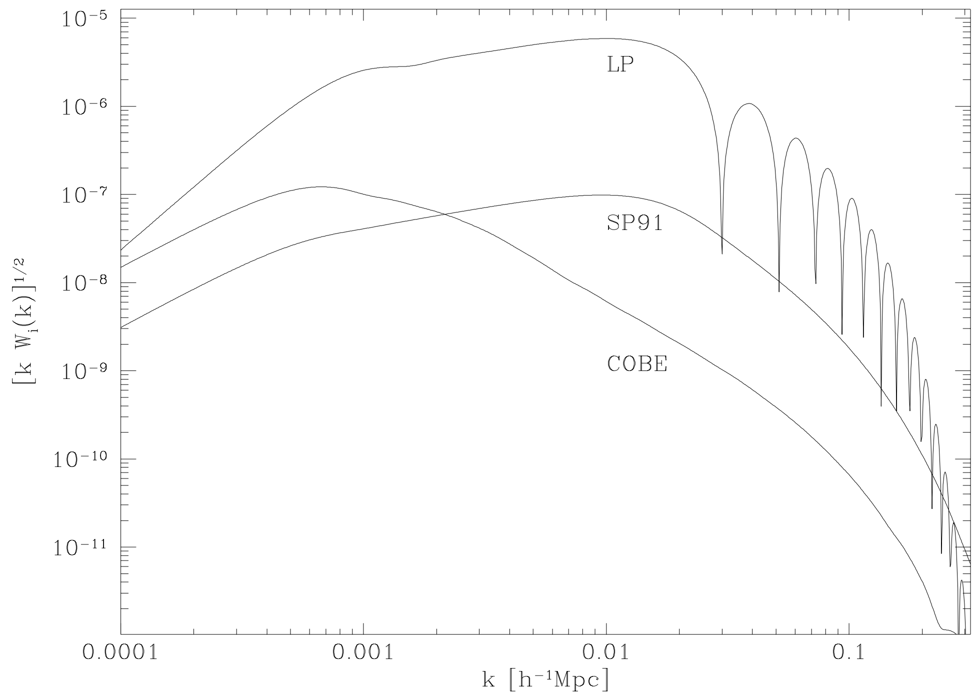


Figure 7.2: Window functions.

The window functions of the Lauer-Postman bulk flow measurement (LP), the South Pole 1991 nine point scan (SP91), and the COBE DMR 10° pixel r.m.s. measurement (COBE) are plotted as a function of comoving wavenumber k.

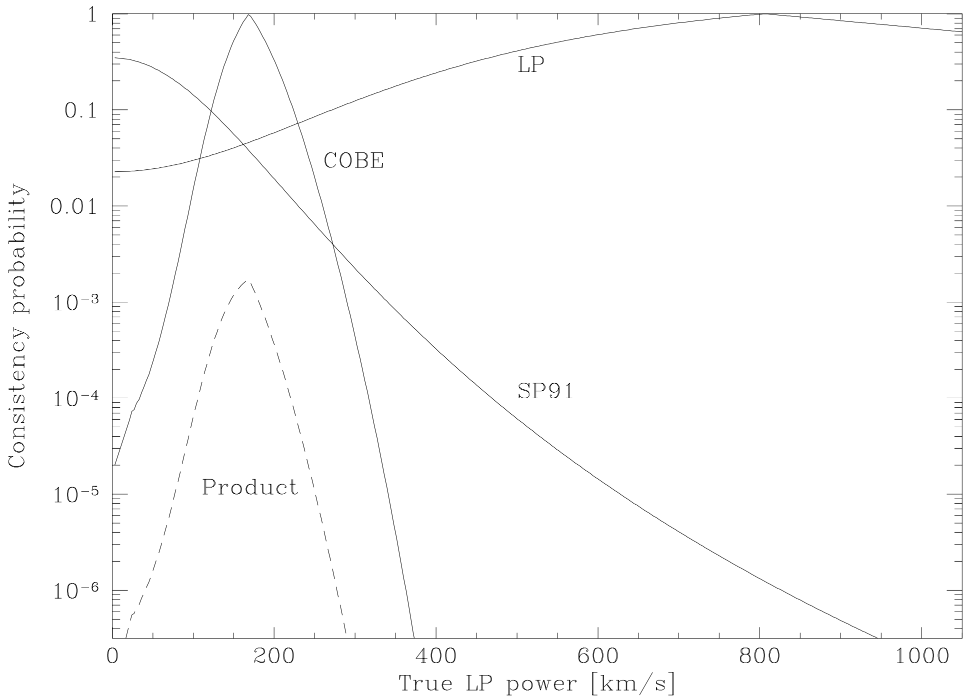


Figure 7.3: Consistency probabilities.

The probability that the the Lauer-Postman bulk flow measurement (LP), the COBE DMR experiment and the South Pole 1991 nine point scan (SP91) are consistent with CDM is plotted as a function of the normalization of the power spectrum. The normalization is expressed in terms of the expected bulk flow in a LP measurement. The dashed line is the product of these three probabilities, and takes a maximum for a normalization corresponding to 168 km/s.

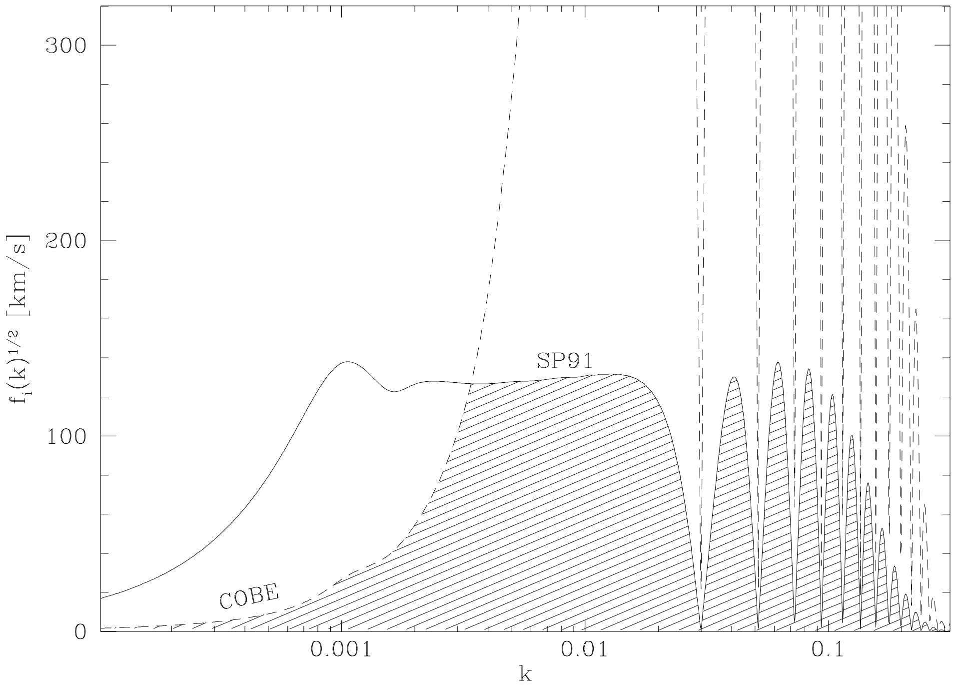


Figure 7.4: The best places to hide power.

The functions f_{sp} (solid line) and f_{cobe} (dashed line) are plotted as a function of wavenumber k. The shaded region, *i.e.* the area lying beneath both curves, constitutes the LP bulk flows that would be consistent with the SP91 and COBE experiments using the $N = 1$ constraints only, when the power spectrum is a single delta function located at k.

Appendix A

The Efficiency Parameter f_{ion}

In this appendix, we will discuss the parameter f_{ion}, and see that it rarely drops below 30%. All the notation is that of Chapter 3. We will first discuss the thermal evolution of intergalactic hydrogen exposed to a strong UV flux, and then use the results to write down a differential equation for the volume fraction of the universe that is ionized, subject to point sources of UV radiation that switch on at different times. We will see that photoionization is so efficient within the ionized regions of the IGM that quite a simple equation can be given for the expansion of the ionized regions.

The evolution of IGM exposed to ionizing radiation has been discussed by many authors. Important early work includes that of Arons & McCray (1970), Bergeron & Salpeter (1970) and Arons & Wingert (1972). The main novelty of the treatment that follows is that whereas previous treatments focus on late ($z < 5$) epochs, when various simplifying approximations can be made because the recombination and Compton rates are low, we are mainly interested in the case $50 < z < 150$. We show that IGM exposed to a strong UV flux rapidly approaches a quasistatic equilibrium state, where it is almost fully ionized and the temperature is such that photoionization heating exactly balances Compton cooling. This simplifies the calculations dramatically, since the entire thermal history of the IGM can be summarized by a single function $\chi(z)$, the volume fraction that is ionized. Thus a fraction $\chi(z)$ is ionized and hot (with a temperature that depends only on z, not on when it became ionized), and a fraction $1-\chi(z)$ is neutral and cold.

In the first section, we justify this approximation. In the second section, we derive a differential equation for the time-evolution of χ as well as a useful analytic estimate of f_{ion}.

A.1 Intergalactic Strömgren Spheres

Let x denote the ionization fraction in a small, homogeneous volume of intergalactic hydrogen, *i.e.* $x \equiv n_{HII}/(n_{HI}+n_{HII})$. ($x$ is not to be confused with χ, the volume fraction in ionized bubbles.) When this IGM is at temperature T, exposed to a density of η UV photons per proton, the ionization fraction x evolves as follows:

$$\frac{dx}{d(-z)} = \frac{1+z}{\sqrt{1+\Omega_0 z}}\left[\lambda_{pi}(1-x)+\lambda_{ci}x(1-x)-\lambda_{rec}^{(1)}x^2\right], \tag{A.1}$$

where $H_0^{-1}(1+z)^{-3}$ times the rates per baryon for photoionization, collisional ionization and recombination are given by

$$\begin{cases} \lambda_{pi} \approx 1.04 \times 10^{12} \left[h\Omega_b\sigma_{18}\right]\eta, \\ \lambda_{ci} \approx 2.03 \times 10^4 h\Omega_b T_4^{1/2} e^{-15.8/T_4,} \\ \lambda_{rec}^{(1)} \approx 0.717 h\Omega_b T_4^{-1/2}\left[1.808 - 0.5\ln T_4 + 0.187 T_4^{1/3}\right], \end{cases} \tag{A.2}$$

and $T_4 \equiv T/10^4$K. σ_{18} is the thermally averaged photoionization cross section in units of 10^{-18}cm^2, and has been computed in Table 3.3 for various spectra using the differential cross section from Osterbrock (1974). The collisional ionization rate is from Black (1981). The recombination rate is the total rate to all hydrogenic levels (Seaton 1959).

Below we will see that in the ionized Strömgren bubbles that will appear around the galaxies or quasars, the photoionization rate is so much greater than the other rates that to a good approximation, equation (A.1) can be replaced by the following simple model for the IGM:

- It is completely ionized ($x = 1$).
- When a neutral hydrogen atom is formed through recombination, it is instantly photoionized again.

Thus the only unknown parameter is the IGM temperature, which determines the recombination rate, which in turn equals the photoionization rate and thus determines the rate of heating.

Let us investigate when this model is valid. Near the perimeter of an ionized Strömgren sphere of radius r surrounding a galaxy, the number of UV photons per proton is roughly

$$\eta = \frac{S_{uv}}{4\pi r^2 cn},$$

where S_{uv} is the rate at which UV photons leave the galaxy. For an O5 star, the photon flux above the Lyman limit is approximately $3.1 \times 10^{49} s^{-1}$ (Spitzer 1968), so if each N solar masses of baryons in a galaxy leads to production of a UV flux equivalent to that of an O5 star, then

$$\eta \geq 0.77 \frac{f_{esc} M_6}{h^2 r_1^2 N(1+z)^3}, \tag{A.3}$$

inside the sphere, where $r_1 \equiv r/1$Mpc and $M_6 \equiv M/10^6 M_\odot$. When a fraction f_s of all matter has formed galaxies of a typical total (baryonic and dark) mass M, then in the absence of strong clustering, the typical separation between two galaxies is

$$R = \left(\frac{M}{f_s\rho}\right)^{1/3} \approx \left(\frac{15\text{kpc}}{1+z}\right)\left(\frac{M_6}{h^2 f_s}\right)^{1/3},$$

where $M_6 \equiv M/10^6 M_\odot$. Thus r continues to increase until $r \approx R$, and spheres from neighboring galaxies begin to overlap. We are interested in the regime where $z < 150$. Substituting this and equation (A.3) into (A.2), we see that $\lambda_{pi} \gg \lambda_{ci}$ and $\lambda_{pi} \gg \lambda_{rec}$ for any reasonable parameter values. Hence we can neglect collisional ionization in equation (A.1). Since $\lambda_{pi} \gg 1$, the photoionization timescale is much shorter than the Hubble timescale, so equation (A.1) will quickly approach a quasistatic equilibrium solution where the recombination rate equals the photoionization rate, *i.e.*

$$x \approx 1 - \frac{\lambda_{rec}}{\lambda_{pi}} \approx 1.$$

In conclusion, the simple $x = 1$ model is valid for all parameter values in our regime of interest.

When a hydrogen atom gets ionized, the photoelectron acquires an average kinetic energy of $\frac{3}{2}kT^*$, where T^* is defined by $\frac{3}{2}kT^* = \langle E_{uv}\rangle - 13.6\text{eV}$, and $\langle E_{uv}\rangle$ is the average energy of the ionizing photons (see Table 3.3).

Since the timescale for Coulomb collisions is much shorter than any other timescales involved, the electrons and protons rapidly thermalize, and we can always assume that their velocity distribution is Maxwellian, corresponding to some well-defined temperature T. Thus shortly after the hydrogen gets photoionized, after the electrons have transferred half of their energy to the protons, the plasma temperature is $T = \frac{1}{2}T^*$.

The net effect of a recombination and subsequent photoionization is to remove the kinetic energy of the captured electron, say $\frac{3}{2}kT\eta_{rec}(T)$, from the gas and replace it with $\frac{3}{2}kT^*$, the kinetic energy of the new photoelectron. Since the recombination cross section is approximately proportional to v^{-2}, slower electrons are more likely to get captured. Hence the mean energy of the captured electrons is slightly lower than $\frac{3}{2}kT$, *i.e.* $\eta_{rec}(T)$ is slightly less than unity (Osterbrock 1974). We compute $\eta_{rec}(T)$ using Seaton (1959). The complication that $\eta_{rec}(T) \neq 1$ turns out to be of only marginal importance: $\eta_{rec}(10^4\text{K}) \approx 0.8$, which only raises the equilibrium temperatures calculated below by a few percent.

The higher the recombination rate, the faster this effect will tend to push the temperature up towards T^*. The two dominant cooling effects are Compton drag against the microwave background photons and cooling due to the adiabatic expansion of the universe. Line cooling from collisional excitations can be neglected, since the neutral fraction $1-x \approx 0$. Combining these effects, we obtain the evolution equation for the IGM inside of a Strömgren bubble:

$$\frac{dT}{d(-z)} = -\frac{2}{1+z}T + \frac{1+z}{\sqrt{1+\Omega_0 z}}\left[\lambda_{comp}(T_{cbr}-T) + \frac{1}{2}\lambda_{rec}(T)[T_{cbr} - \eta_{rec}(T)T]\right] \tag{A.4}$$

where

$$\lambda_{comp} = \frac{4\pi^2}{45}\left(\frac{kT_{cbr}}{\hbar c}\right)^4 \frac{\hbar\sigma_t}{H_0 m_e}(1+z)^{-3} \approx 0.00418h^{-1}(1+z)$$

is $(1+z)^{-3}$ times the Compton cooling rate per Hubble time and $T_{cbr} = T_{cbr,0}(1+z)$. We have taken $T_{cbr,0} \approx 2.726\text{K}$ (Mather *et al.* 1994). The factor of $\frac{1}{2}$ in front of the λ_{rec} term is due to the fact that the photoelectrons share their acquired energy with the protons. The average energy of the ionizing photons is given by the spectrum $P(\nu)$ as $\langle E_{uv}\rangle = h\langle\nu\rangle$, where

$$\langle\nu\rangle = \frac{\int_0^\infty P(\nu)\sigma(\nu)d\nu}{\int_0^\infty \nu^{-1}P(\nu)\sigma(\nu)d\nu}.$$

Here the photoionization cross section $\sigma(\nu)$ is given by Osterbrock (1974). Note that, in contrast to certain nebula calculations where all photons get absorbed sooner or later, the spectrum should be weighted by the photoionization cross section. This is because most photons never get absorbed in the Strömgren regions (only in the transition layer), and all that is relevant is the energy distribution of those photons that do. $P(\nu)$ is the energy distribution (W/Hz), not the number distribution which is proportional to $P(\nu)/\nu$.

The spectral parameters $\langle E_{uv}\rangle$ and T^* are given in Table 3.3 for some selected spectra. A Planck spectrum $P(\nu) \propto \nu^3/\left(e^{h\nu/kT}-1\right)$ gives quite a good prediction of T^* for stars with surface temperatures below 30,000K. For very hot stars, more realistic spectra (Vacca, 1993) have a sharp break at the Lyman limit, and fall off much slower above it, thus giving higher values of T^*. As seen in Table 3.3, an extremely metal poor star of surface temperature 50,000K gives roughly the same T^* as QSO radiation. The only stars that are likely to be relevant to early photoionization scenarios are hot and short-lived ones, since the universe is only about 10^7 years old at $z = 100$, and fainter stars would be unable to inject enough energy in so short a time. Conceivably, less massive stars could play a dominant role later on, thus lowering T^*. However, since they radiate such a small fraction of their energy above the Lyman limit, very large numbers would be needed, which could be difficult to reconcile with the absence of observations of Population III stars today. If black holes are the dominant UV source, the stellar spectra of Table 3.3 are obviously irrelevant. A power law spectrum $P(\nu) \propto \nu^{-\alpha}$ with $\alpha = 1$ fits observed QSO spectra rather well in the vicinity of the Lyman limit (Cheney & Rowan-Robinson 1981; O'Brien *et al.* 1988), and is also consistent with the standard model for black hole accretion.

Numerical solutions to equation (A.4) are shown in Figure 3.9, and it is seen that the temperature evolution separates into three distinct phases. In the first phase, the IGM is outside of the Strömgren regions, unexposed to UV radiation, and remains cold and neutral. In the second phase, the IGM suddenly becomes ionized, and its temperature instantly rises to $\frac{1}{2}T^*$. After this, Compton cooling rapidly reduces the temperature to a quasi-equilibrium value of a few thousand K. After this, in the third phase, T changes only quite slowly, and is approximately given by setting the expression in square brackets in equation (A.4) equal to zero. This quasi-equilibrium temperature is typically many times lower than T^*, since Compton cooling is so efficient at the high redshifts involved.

A.2 The Expansion of Strömgren Regions

This rapid approach to quasi-equilibrium, where the IGM "loses its memory" of how long ago it became part of a Strömgren region, enables us to construct a very simple model for the ionization history of the universe. At redshift z, a volume fraction $\chi(z)$ of the universe is completely ionized and typically has a temperature of a few thousand K. The ionized part need not consist of non-overlapping spheres; it can have any topology whatsoever. The remainder is cold and neutral.

Between the ionized and neutral regions is a relatively thin transition layer, where the IGM becomes photoionized and its temperature adjusts to the quasistatic value as in Figure 3.9. As this IGM becomes part of the hot and ionized volume, the transition layer moves, and $\chi(z)$ increases[1].

[1] We are tacitly assuming that the UV luminosity of the galaxy that creates each Strömgren sphere never

As long as $\chi < 1$, all UV photons produced are absorbed instantly to a good approximation. Thus the rate at which UV photons are released is the sum of the rate at which they are used to counterbalance recombinations inside the hot bubbles and the rate at which they are used to break new ground, to increase χ. Thus

$$f_{uvpp}\frac{df_s}{dt} = \alpha^{(2)}(T)n\chi + \frac{d\chi}{dt},$$

where $\alpha^{(2)}(T)$ is the total recombination rate to all hydrogenic levels except the ground state[2]. Changing the independent variable to redshift and using equation (3.6), we find that this becomes

$$\frac{d\chi}{d(-z)} + \lambda_{rec}^{(2)}\frac{1+z}{\sqrt{1+\Omega_0 z}}\chi = \frac{2}{\sqrt{\pi}}\left(\frac{f_{uvpp}}{1+z_{vir}}\right)\exp\left[-\left(\frac{1+z}{1+z_{vir}}\right)^2\right]. \tag{A.5}$$

Here

$$\lambda_{rec}^{(2)} \approx 0.717h\Omega_b T_4^{-1/2}\left[1.04 - 0.5\ln T_4 + 0.19T_4^{1/2}\right]$$

is $H_0^{-1}(1+z)^{-3}$ times the total recombination rate per baryon to all hydrogenic levels except the ground state. The fit is to the data of Spitzer (1968) and is accurate to within 2% for $30\text{K} < T < 64,000\text{K}$. $\lambda_{rec}^{(2)}$ is to be evaluated at the quasi-equilibrium temperature $T(z)$ discussed above.

Using the values in Table 3.3 for the pessimistic, middle-of-the-road and optimistic estimates, the parameter f_{uvpp} equals roughly 4, 190 and 24,000, respectively.

In the absence of photon waste through recombination, equation equation (A.5) would have the solution $\chi^*(z) = f_{uvpp}f_s(z)$, so the ionization efficiency is

$$f_{ion}(z) = \chi(z)/\chi^*(z).$$

Since equation (A.5) is linear in χ and the initial data is $\chi = 0$ at some redshift, it is readily seen that the solution $\chi(z)$ is proportional to f_{uvpp}, the constant in front of the source term. Combining these last two observations, we see that f_{ion} is independent of f_{uvpp} and hence independent of the poorly known parameters f_{burn}, f_{uv} and f_{esc}.

Plots of $f_{ion}(z)$ from numerical solutions of equation (A.5) are shown in Figure 3.10 for various parameter values, and it is seen that the dependence on z is generally quite weak. Let us make use of this fact by substituting the *Ansatz* $\chi(z) = f_{ion}(z)f_{uvpp}f_s(z)$ into equation (A.5), and setting $f_{ion}'(z) \approx 0$. Using equation (3.6) and an asymptotic

decreases. Although obviously untrue, this is in fact an excellent approximation, since these early dwarf-galaxies correspond to perturbations far out in the Gaussian tail. Since $f_s(z)$ grows so dramatically as the redshift decreases and we move from five sigma to four sigma to three sigma, *etc.*, almost all galaxies in existence at a given redshift are in fact very young, so that older ones that have begun to dim can be safely neglected.

[2] A reionization directly to the ground state produces a UV photon that usually propagates uninterrupted through the highly ionized Strömgren region, and then ionizes another atom in the transition layer between the expanding Strömgren region and its cold and neutral surrounding. Thus recombinations directly to the ground state were included in the above calculation of the quasi-equilibrium temperature of the Strömgren bubbles, since the resulting UV photons could be considered lost from the latter. Here, on the contrary, recombinations directly to the ground state should not be included, since the UV photons they produce are not wasted from an energetics point of view.

approximation for the error function, we obtain

$$f_{ion}(z) \approx \frac{1}{1 + 0.48\lambda_{rec}^{(2)}(1+z_{vir})^2/\sqrt{1+\Omega_0 z}},$$

independent of f_{uvpp}, which agrees to within 10% with the numerical solutions for all reasonable parameter values. This expression highlights the connection between f_{ion} and the thermal evolution of the Strömgren bubbles: Essentially, the higher the quasi-static temperature, the lower the recombination rate $\lambda_{rec}^{(2)}$, and the higher f_{ion} becomes.

The value of f_{ion} relevant to computing the ionization redshift is obviously that where $z = z_{ion}$. As we have seen, z_{ion} typically lies between $2z_{vir}$ and $3z_{vir}$. Substituting $T \approx 2,500$K into the expression for λ_{rec} and taking $\Omega_0 \approx 1$ and $z = z_{ion} \approx 2.5z_{vir}$, the above reduces to

$$f_{ion} \approx \frac{1}{1 + 0.8h\Omega_b(1+z_{vir})^{3/2}},$$

so we see that f_{ion} will be of order unity unless $z_{vir} \gg 15$ or $h\Omega_b \gg 0.03$.

Appendix B

Comparing Goodness-of-fit Parameters

In this appendix, we compare the performances of the probability product and the likelihood product as goodness-of-fit parameters. The notation is that of Chapter 7.

First of all, what do we mean by a goodness-of-fit parameter η being good? That η leads to the correct model being ruled out at some confidence level x cannot be held against it — by definition of significance level, this happens a fraction $(1-x)$ of the time. Rather, the conventional criterion for rating goodness-of-fit parameters is *rejection power*: given a model and a set of observations, one η is said to be more powerful than another if it rejects the model at a higher level of significance. An example of a very stupid goodness-of-fit parameter, which in fact has the lowest rejection power possible, is a random variable η drawn from a uniform distribution on $[0,1]$, thus containing no information whatsoever about the model or the observed data. Use of this parameter will reject the model at 95% confidence only 5% of the time, even if the data is blatantly inconsistent with the model.

B.1 Both η_l and η_p can be "fooled"...

Given a random variable s_i with probability distribution f_i, we define the corresponding random variable for likelihood by $L_i = f_i(s_i)/f_{max}$, where we chose the normalization constant $f_{max} \equiv \max_x f_i(x)$ so that we will always have $0 \leq L_i \leq 1$. Thus for N experiments, the likelihood product

$$\eta_l \equiv \prod_{i=1}^{N} L_i$$

will also be a random variable on the interval $[0,1]$.

It is easy to construct examples where either the probability product η_p (as defined in Section 7.2.1) or the likelihood product η_l give very low rejection power. The Achilles' heel of the probability product is multimodal distributions, where values near the mean are rather unlikely. For example, suppose $N = 1$ and we have a double-humped distribution such as

$$f_1(s) \propto s^2 e^{-s^2}.$$

If we observe a value $\hat{s}_1 = 0$, then η_l would reject the model with 100% confidence whereas the probability product fails miserably, rejecting with 0% confidence since $\hat{s}_1$ equals the mean. The likelihood product, on the other hand, has the weakness that the highest likelihood may be attained far out in the tail of the distribution. Suppose for instance that $N = 5$ and we have triangle distributions

$$f_i(s) = 2s$$

on the interval $0 \leq s \leq 1$. If all five observed values $\hat{s}_i$ lie between 0.999 and 1, something is clearly wrong with the model. Since $\hat{\eta}_p > 0.004^5$, using equation (7.1) indeed rejects the model at a confidence level exceeding $1 - F_\eta(0.004^5) \approx 99.999997\%$. Yet the likelihood product is near its maximum value: $\hat{\eta}_l \geq 0.999^5 \approx 0.995$, which gives virtually no rejection power at all.

B.2 ...but they usually give similar results.

It is important to note that neither of the two examples above were particularly physical. The random variables arising from cosmic variance have Gaussian or chi-squared distributions, and the same tends to holds for the various experimental noise distributions. Thus the probability distributions to which our goodness-of-fit parameter is applied in this chapter are unimodal (which eliminates the first example) and taper off to zero smoothly (which eliminates the second example). Hence for cosmological applications, goodness-of-fit parameters should not be rated by their performance with such pathological distributions, but rather by their rejection power when applied to continuous, unimodal distributions. We will now compare the performance of η_p and η_l for a few such cases.

For a symmetric *exponential* distribution

$$f_i(s) = \frac{1}{2}e^{-|s|},$$

it is easy to see that the likelihood L_i has a uniform distribution. This means that η_p and η_l will have identical distributions, for arbitrary N. It is straightforward to show that the same holds for symmetric *triangle* distributions

$$f_i(s) = 1 - |s|.$$

A third case where η_p and η_l give identical results is when $N = 1$ and f is any smooth unimodal function.

For a *Gaussian* distribution

$$f_i(s) = (2\pi)^{-1/2}e^{-s^2/2},$$

we have

$$P(L_i < x) = 2\mathrm{erfc}\left[(-2\ln x)^{1/2}\right].$$

Although the probability distribution of η_l appears not to be expressible in terms of elementary functions for arbitrary n, it is easy to show that η_l has a uniform distribution for the special case $N = 2$. Thus the likelihood product gives rejection at a confidence level of

$$P(\eta_l < \hat{\eta}_l) = e^{-(\hat{s}_1^2+\hat{s}_2^2)/2}.$$

Comparing this with the corresponding confidence level based on η_p shows a remarkable agreement between the two methods. Within the disc $\hat{s}_1^2 + \hat{s}_2^2 < 4$, over which $P(\eta < \hat{\eta})$ varies with many orders of magnitude, the two methods never differ by more than a factor of two. Thus, at worst, one may yield a confidence level of say 99.98% where the other yields 99.99%. The probability product is stronger in slightly more than half of the $(\hat{s}_1, \hat{s}_2)$-plane, roughly for regions that are more than 20° away from any of the coordinate axes.

In conclusion, we have seen that for unimodal, continuous probability distributions, the likelihood product and the probability product tend to give fairly similar — in a few special cases even identical — results. Thus chosing one parameter over the other is more a matter of personal preference than something that is likely to seriously affect any scientific conclusions. There is however one important practical consideration: The probability distribution of η_l depends on the probability distributions of the random variables s_i. This means that, apart from a few fortuitous special cases such as described above, it can generally not be calculated analytically. Rather, it must be computed numerically, through numerical convolution or Monte Carlo simulation. The probability distribution for η_p, on the other hand, is always known analytically, as given by equation (7.1), so the probability product is considerably simpler to use.

Appendix C

Window Functions

All the notation in this appendix is defined in Chapter 7. The results of CMB anisotropy experiments can be conveniently described by expanding the temperature fluctuation in spherical harmonics:

$$\frac{\Delta T}{T}(\hat{\mathbf{r}}) = \sum_{l=2}^{\infty} \sum_{m=-l}^{l} a_{lm} Y_{lm}(\hat{\mathbf{r}}).$$

(The monopole and dipole anisotropies have been removed from the above expression, since they are unmeasurable.) If the fluctuations are Gaussian, then each coefficient a_{lm} is an independent Gaussian random variable with zero mean (Bond & Efstathiou 1987). The statistical properties of the fluctuations are then completely specified by the variances of these quantities

$$C_l \equiv \left\langle |a_{lm}|^2 \right\rangle.$$

(The fact that the variances are independent of m is an immediate consequence of spherical symmetry.) Different CMB experiments are sensitive to different linear combinations of the C_l's:

$$S = \sum_{l=2}^{\infty} F_l C_l, \tag{C.1}$$

where S is the ensemble-averaged mean-square signal in a particular experiment, and the "filter function" F_l specifies the sensitivity of the experiment on different angular scales. The filter functions for COBE and SP91 are

$$\begin{cases} F_l^{(cobe)} & = \frac{(2l+1)}{4\pi} e^{-\sigma_c^2 (l+\frac{1}{2})^2}, \\ F_l^{(sp)} & = 4 e^{-\sigma_s^2 (l+\frac{1}{2})^2} \sum_{m=-l}^{l} H_0^2(\alpha m), \end{cases} \tag{C.2}$$

where H_0 is a Struve function. $\sigma_c = 4.25^\circ$ and $\sigma_s = 0.70^\circ$ are the r.m.s. beamwidths for the two experiments, and $\alpha = 1.5^\circ$ is the amplitude of the beam chop (Bond *et al.* 1991; Dodelson & Jubas 1993; White *et al.* 1993).

For Sachs-Wolfe fluctuations in a spatially flat Universe with the standard ionization history, the angular power spectrum C_l is related to the power spectrum of the matter fluctuations in the following way (Peebles 1984; Bond & Efstathiou 1987):

$$C_l = \frac{8}{\pi \tau_0^4} \int_0^{\infty} dk P(k) \bar{j}_l^2(k). \tag{C.3}$$

Here τ_0 is the conformal time at the present epoch, and

$$\bar{j}_l(k) \equiv \int j_l(k\tau)V(\tau)\,d\tau,$$

where j_l is a spherical Bessel function. The visibility function V is the probability distribution for the conformal time at which a random CMB photon was last scattered. $\bar{j}_l(k)$ is therefore the average of $j_l(k\tau)$ over the last scattering surface. We have used the V of Padmanabhan (1993).

We can combine equations (C.1), (C.2), and (C.3) to get the window functions for the two experiments:

$$\begin{cases} W_{cobe} & = \frac{2}{\pi^2 k^2 \tau_0^4} \sum_{l=2}^{\infty} \bar{j}_l^2(k) e^{-\sigma_c^2 \left(l+\frac{1}{2}\right)^2} (2l+1) \\ W_{sp} & = \frac{32}{\pi k^2 \tau_0^4} \sum_{l=2}^{\infty} \bar{j}_l^2(k) e^{-\sigma_s^2 \left(l+\frac{1}{2}\right)^2} \sum_{m=-l}^{l} H_0^2(\alpha m) \end{cases}$$

The mean-square bulk flow inside of a sphere of radius a is (see, *e.g.*, Kolb & Turner 1990)

$$\left\langle v^2 \right\rangle = \int dk P(k) \frac{18}{\pi^2 \tau_0^2} \frac{j_1^2(ka)}{(ka)^2}. \tag{C.4}$$

However, we must make two corrections to this result before applying it to the LP data. This formula applies to a measurement of the bulk flow within a sphere with an infinitely sharp boundary. In reality, errors in measuring distances cause the boundary of the spherical region to be somewhat fuzzy. If we assume that distance measurements are subject to a fractional error ϵ, then the window function must be multiplied by $e^{-(\epsilon ka)^2}$. We have taken $\epsilon = 0.16$, the average value quoted by LP. It should be noted that this value varies from galaxy to galaxy in the LP sample, due to the distance estimation technique used, and that a more accurate window function that reflects the discrete locations of the Abell clusters used in the survey should take this into account.

The second correction has to do with the behavior of the window function at small k. Equation (C.4) applies to the velocity relative to the rest frame of the Universe. The velocity measured by LP is with respect to the CMB rest frame. If there is an intrinsic CMB dipole anisotropy, then these two reference frames differ. Therefore, we must include in equation (C.1), a term corresponding to the intrinsic CMB dipole. This correction was first noticed by Górski (1991). After applying both of these corrections, the LP window function is

$$W_{lp} = \frac{18}{\pi^2 \tau_0^2} \left(\frac{j_1(ka)}{ka} e^{-(\epsilon ka)^2} - \frac{\bar{j}_1(k\tau_0)}{k\tau_0} \right)^2.$$

References

Alcock, C. *et al.* 1993, *Nature*, **365**, 621.

Alpher, R. A., Bethe, H. A. & Gamow, G. 1948, *Phys. Rev.*, **73**, 803.

Anninos, P. *et al.* 1991, *ApJ*, **382**, 71.

Arnaud, M. *et al.* 1992, *A&A*, **254**, 49.

Arons. J. & McCray, R. 1970, *Ap. Lett.*, **5**, 123.

Arons, J. & Wingert, D. W. 1972, *ApJ*, **177**, 1.

Bardeen, J. M., Bond, J. R., Kaiser, N. & Szalay, A. S. 1986, *ApJ*, **304**, 15.

Bartlett, J. & Stebbins, A. 1991, *ApJ*, **371**, 8.

Bergeron, J. & Salpeter, E. E. 1970, *Ap. Lett.*, **7**, 115.

de Bernardis, P. *et al.* 1994, *ApJ*, **422**, L33.

Binggeli, B., Sandage, A. & Tammann, G. A. 1988, *ARA&A*, **26**, 509.

Binney, J. 1977, *ApJ*, **215**, 483.

Birkinshaw, M. & Hughes, J. P. 1994, *ApJ*, **420**, 33.

Black, J. 1981, *MNRAS*, **197**, 553.

Blanchard, A., Valls-Gabaud, D. & Mamon, G. A. 1992, *A&A*, **264**, 365.

Blumenthal, G. R., Faber S. M., Primack, J. R., & Rees, M. J.1984, *Nature*, **311**, 517.

Bond, J. R. & Efstathiou, G. 1984, *ApJ*, **285**, L45.

Bond, J. R. & Szalay, A. S. 1983, *ApJ*, **274**, 443.

Brainerd, T. G. & Villumsen, J. V. 1992, *ApJ*, **394**, 409.

Bruhweiler, F. C., Gull, T. R., Kafatos, M., & Sofia, S. 1980, *ApJ*, **238**, L27.

Carlberg, R. G. & Couchman, H. M. P. 1989, *ApJ*, **340**, 47.

Carroll, S. M., Press, W. H. & Turner, E. L. 1992, *ARA&A*, **30**, 499.

Cen, R., Gnedin, N. Y., Koffmann, L. A., & Ostriker, J. P. 1992, *ApJ*, **399**, L11.

Cen, R., Ostriker, J. P. & Peebles, P. J. E. 1993, *ApJ*, **415**, 423.

Cheney, J. E. & Rowan-Robinson, M. 1981, *MNRAS*, **195**, 831.

Cheng, E. *et al.* 1994, *ApJ*, **422**, L37.

Cioffi, D. F., McKee, C. F., & Bertschinger, E. 1988, *ApJ*, **334**, 252.

Conklin, E. K. 1969, *Nature*, **222**, 971.

Couchman, H. M. P. 1985, *MNRAS*, **214**, 137.

Couchman, H. M. P. & Rees, M. J. 1986, *MNRAS*, **221**, 53.

Cox, D. P. & Smith, B. W. 1974, *ApJ*, **189**, L105.

Dalgarno, A. & McCray, R. A. 1972, *ARA&A*, **10**, 375.

David, L. P., Arnaud, K. A., Forman, W. & Jones, C. 1990, *ApJ*, **356**, 32.

David, L. P., Forman, W. & Jones, C. 1991, *ApJ*, **369**, 121.

Davis, M., Summers, F. J. & Schlegel, D. 1992, *Nature*, **359**, 393.

Dekel, A. & Silk, J. 1986, *ApJ*, **303**, 39.

Devlin, M. *et al.* 1992, in *Proc. NAS Colloquium on Physical Cosmology*, Physics Reports (in press).

Dicke, R. H., Peebles, P. J. E., Roll, P. G. & Wilkinson, D. T. 1965, *ApJ*, **142**, 414.

Dodelson, S., Gyuk, G. & Turner, M. S. 1994a, *Phys. Rev. Lett.*, **72**, 3754.

Dodelson, S., Gyuk, G. & Turner, M. S. 1994b, *Phys. Rev. D*, **49**, 5068.

Dodelson, S. & Jubas, J.M. 1993, *Phys. Rev. Lett.*, **70**, 2224.

Dolgov, A. D., Sazhin, M. V. & Zel'dovich, Y. B. 1990, *Modern Cosmology*, Editions Frontieres, Gif-sur-Yvette, France.

Donahue, M. & Shull, J. M. 1991, *ApJ*, **383**, 511.

Dragovan, M. *et al.* 1993, preprint.

Edge, A. C. 1989, Ph.D. thesis, University of Leicester.

Efstathiou, G., Bond, J. R. & White, S. D. M. 1992, *MNRAS*, **258**, 1P.

Efstathiou, G., Frenk, C. S., White, S. D. M., & Davis, M. 1985, *ApJ*, **57**, 241.

Efstathiou, G., Frenk, C. S., White, S. D. M., & Davis, M. 1988, *MNRAS*, **235**, 715.

Efstathiou, G. & Rees, M. J. 1988, *MNRAS*, **230**, 5P.

Feldman, H. A. & Watkins, R. 1993, preprint.

Feynman, R. P. 1939, *Phys. Rev.*, **56**, 340.

Gamow, G. 1970, *My World Line*, Viking, New York.

Gaier, T. *et al.* 1992, *ApJ*, **398**, L1.

Gelb, J. M. & Bertschinger, E. 1992, preprint.

Ginsburg, V. L. & Ozernoi, L. M. 1965, *Astron. Zh.*, **42**, 943.
(Engl. transl. 196 Sov. Astron. AJ, 9, 726)

Glanfield, J. R. 1966, *MNRAS*, **131**, 271.

Gnedin, N. Y. & Ostriker, J. P.1992, *ApJ*, **400**, 1.

Górski, K. M. 1991, *ApJ*, **370**, L5.

Górski, K. M. 1992, *ApJ*, **398**, L5.

Górski, K. M., Stompor, R. & Juszkiewicz, R. 1993, *ApJ*, **410**, L1.

Gott, J. R. & Rees, M. J. 1975, *A&A*, **45**, 365.

Gundersen, J.O. *et al.* 1993, *ApJ*, **413**, L1.

Gunn, J. E. & Peterson, B. A. 1965, *ApJ*, **142**, 1633.

Gyuk, G. & Turner, M. S. 1994, astro-ph/9403054 preprint.

Hancock, S. *et al.* 1994;Nature;367;333

Hatsukade, I. 1989, thesis, Osaka University.

Heckman, T. M., Armus, L., & Miley, G. K. 1990, *ApJS*, **74**, 833.

Henry, P. S. 1971, *Nature*, **231**, 516.

Holtzman, J. A. 1989, *ApJS*, **71**, 1.

Hu, W., Scott, D. & Silk, J. 1994, *Phys. Rev. D*, **49**, 648.

Hu, W. & Silk, J. 1993, *Phys. Rev. D*, **48**, 485.

Hubble, E. 1929, *Proc. N. A. S.*, **15**, 168.

Hughes, J. P., Yamashita, K., Okumura, Y., Tsunemi, H., & Matsuoka, M. 1988, *ApJ*, **327**, 615.

Ikeuchi, S. 1981, *Publ. Astr. Soc. Jpn.*, **33**, 211.

Ikeuchi, S. & Ostriker, J. P. 1986, *ApJ*, **301**, 522.

Ikeuchi, S. & Turner, E. L 1991, *ApJ*, **381**, L1.

Jaffe, A. H., Stebbins, A. & Frieman, J. A. 1994, *ApJ*, **420**, 9.

Jubas, J. & Dodelson, S. 1993, preprint.

Juszkiewicz, R. 1993, private communication.

Juszkiewicz, R., Górski, K. M., & Silk, J.1987, *ApJ*, **323**, L1.

Kamionkowski, M. & Spergel, D. N. 1994, *ApJ*, **432**, 7.

Kamionkowski, M., Spergel, D. N. & Sugiyama, N. 1994, *ApJ*, **426**, L57.

Kashlinsky 1992, *ApJ*, **399**, L1.

Klypin, A., Holtzman, J., primack, J. & Regös, E. 1993, *ApJ*, **416**, 1.

Kolb, E. W. & Turner, M. S. 1990, *The Early Universe*, Addison-Wesley.

Kompaneets, A. S. 1957, *Soviet Phys.-JETP*, **4**, 730.

Lauer, T. & Postman, M. 1994, *ApJ*, **425**, 418.

Lea, S. M., Mushotzky, R., & Holt, S. 1982, *ApJ*, **262**, 24.

Maddox, S. J., Efstathiou, G., Sutherland, W. J. & Loveday, J. 1990, *MNRAS*, **242**, 43.

Malaney, R. A. & Mathews, G. J. 1993, *Phys. Rep.*, **229**, 145.

Mather J. C. *et al.* 1990, *ApJ*, **354**, L37.

Mather *et al.* 1994, *ApJ*, **420**, 439.

Matsumoto, T. *et al.* 1988, *ApJ*, **329**, 567.

McCray, R. & Kafatos, M. 1987, *ApJ*, **317**, 190.

McCray, R. & Snow, T. P. Jr. 1979, *ARA&A*, **17**, 213.

McKee, C. F. & Ostriker, J. P. 1977, *ApJ*, **218**, 148.

Meinhold, P. R. & Lubin, P. M. 1991, *ApJ*, **370**, L11.

Meinhold, P. R. *et al.* 1993, *ApJ*, **409**, L1.

Michelson, A. A. & Morley, E. W. 1887, *Am. Journ. of Science*, **34**, 333.

Miralda-Escudé, J. & Ostriker, J. P. 1990, *ApJ*, **350**, 1.

Mushotzky, R. F. 1984, *Physica Scripta*, **T7**, 157.

Narlikar, J. V. 1993, *Introduction to Cosmology* (2nd ed.), Cambridge U. P., Cambridge.

O'Brien P.T., Wilson, R & Gondhalekar, P. M 1988, *MNRAS*, **233**, 801.

Osterbrock, D. E. 1974, *Astrophysics of Gaseous Nebulae*, Freeman, San Francisco.

Ostriker, J. P. 1991, *Development of Large-Scale Structure in the Universe*, Fermi Lecture Series, Cambridge U. P., Cambridge.

Ostriker, J. P. & Cowie, C. F. 1981, *ApJ*, **243**, L127.

Ostriker, J. P. & McKee, C. F. 1988, *Rev. Mod. Phys.*, **60**, 1.

Oukbir, J. & Blanchard, A. 1992, *A&A*, **262**, L21.

Padmanabhan, T. 1993, *Structure Formation in the Universe*, Cambridge U. P., Cambridge.

Peacock, J. A. & Heavens, A. F. 1990, *MNRAS*, **243**, 133.

Peacock, J. A., Heavens, A. E. & Davies, A. T. 1990 (eds.), *Physics of the Early Universe*, Scottish Universities' Summer School in Physics, IOP Publishing, Bristol.

Peacock, J. A. & Dodds, S. J. 1994, *MNRAS*, **267**, 1020.

Peebles, P. J. E. 1971, *Physical Cosmology*, Princeton U. P., Princeton.

Peebles, P. J. E. 1980, *The large-scale structure of the universe*, Princeton U. P., Princeton.

Peebles, P. J. E. 1984, *ApJ*, **284**, 439.

Peebles, P. J. E. 1987, *ApJ*, **315**, L73.

Peebles, P. J. E 1993, *Principles of Physical Cosmology*, Princeton U. P., Princeton.

Penzias, A. A. & Wilson, R. W. 1965, *ApJ*, **142**, 419.

Pettini, M., Boksenberg, A., & Hunstead, R. 1990, *ApJ*, **348**, 48.

Press, W. H. & Schechter, P. 1974, *ApJ*, **187**, 425.

Rees, M. J. & Ostriker, J. P. 1977, *MNRAS*, **179**, 541.

Rees, M. J.1986, *MNRAS*, **218**, 25P.

Ressell, M. T. & Turner, M. S. 1990, *Comments on Astroph.*, **14**, 323.

Rothenflug, R. L., Vigroux, R., Mushotzky, R., & Holt, S. 1984, *ApJ*, **279**, 53.

Rubin, V. C. & Coyne, G. V. 1988 (eds.), *Large-scale motions in the universe : a Vatican study week*, Princeton U. P., Princeton.

Sadoulet, B. & Cronin, J. W. 1991, *Physics Today*, **44**, 53.

Sanchez, F., Collados, M. & Rebolo, R. (eds.) 1990, *Observational and Physical Cosmology*, 2nd Canary Islands Winter School on Astrophysics, Cambridge U. P., Cambridge.

Schlegel, D., Davis, M., Summers, F. & Holtzman, J. 1993, preprint.

Schuster, J. *et al.*, private communication.

Schwartz, J., Ostriker, J. P., & Yahil, A. 1975, *ApJ*, **202**, 1.

Seaton, M. 1959, *MNRAS*, **119**, 84.

Sedov, L. I. 1959, *Similarity and dimensional methods in mechanics*, Academic, New York.

Shaefer, R. K. & Shafi, Q. 1992, *Nature*, **359**, 199.

Shafi, Q. & Stecker, F. W. 1984, *Phys. Rev. Lett.*, **53**, 1292.

Shapiro, P. R. 1986, *PASP*, **98**, 1014.

Shapiro, P. R. & Giroux, M. L. 1987, *ApJ*, **321**, L107.

Shapiro, H. S. & Tegmark, M. 1994, *Soc. Ind. App. Math. Rev.*, **36**.

Sherman, R. D. 1980, *ApJ*, **237**, 355.

Shuster, J. *et al.* 1993, private communication.

Silk, J. 1977, *ApJ*, **211**, 638.

Smith, M. S., Kawano, L. H. & Malaney, R. A. 1993, *ApJS*, **85**, 219.

Smoot, G. F., Gorenstein, M. V. & Muller, R. A. 1977, *Phys. Rev. Lett.*, **39**, 898.

Smoot, G. F. *et al.* 1992, *ApJ*, **396**, L1.

Spitzer, L. 1968, *Diffuse Matter in Space*, Wiley, New York.

Stebbins, A. & Silk, J. 1986, *ApJ*, **300**, 1.

Steidel, C. C. 1990, *ApJS*, **74**, 37.

Steidel, C. C. & Sargent, W. L. W. 1987, *ApJ*, **318**, L11.

Strauss, M., Cen, R. & Ostriker, J.P. 1993, preprint.

Subrahmanyan, R. *et al.* 1993, *MNRAS*, **263**, 416.

Sugiyama, N. 1993, private communication.

Sugiyama, N., Silk, J. & Vittorio, N. 1993, *ApJ*, **419**, L1.

Suto, Y., Górski, K. M. Juszkiewicz, R., & Silk, J.1988, *Nature*, **332**, 328.

Tegmark, M. 1993, *Found. Phys. Lett.*, **6**, 571.

Tegmark, M., Bunn, E. & Hu, W. 1994, *ApJ*, **434**, 1.

Tegmark, M. & Shapiro, H. S. 1994, *Phys. Rev. E*, **50**, 2538.

Tegmark, M. & Silk, J. 1994, *ApJ*, **423**, 529; **434**, 395

Tegmark, M., Silk, J. & Blanchard, A. 1994, *ApJ*, **420**, 484; **434**, 395

Tegmark, M., Silk, J. & Evrard, A. 1993, *ApJ*, **417**, 54.

Tegmark, M. & Yeh, L. W. 1994, *Physica A*, **202**, 342.

Teresawa, N. 1992, preprint.

Tomisaka, K., Habe, H, & Ikeuchi, S. 1980, *Progr. Theor. Phys. (Japan)*, **64**, 1587.

Vacca, W. 1993, private communication.

Vishniac, E. 1987, *ApJ*, **322**, 597.

Vittorio, N. & Silk, J. 1984, *ApJ*, **285**, L39.

Vittorio, N. & Silk, J. 1992, *ApJ*, **385**, L9.

Walker, P. N. *et al.* 1991, *ApJ*, **376**, 51.

Watson, R.A.,*et al.* 1992, *Nature*, **357**, 660.

Watson, R.A. & Gutiérrez de la Cruz, C.M. 1993, *ApJ*, **419**, L5.

Webb, J. K., Barcons, X., Carswell, R. F., & Parnell, H. C. 1992, *MNRAS*, **255**, 319.

Weaver, R., McCray, R, Castor, J., Shapiro, P., & Moore, R. 1977, *ApJ*, **218**, 377.

Weinberg, S. 1972, *Gravitation and Cosmology*, Wiley, New York.

White, M., Krauss, L., & Silk, J. 1993, *ApJ*, **418**, 535.

White, S. D. M. & Rees, M. J. 1978, *MNRAS*, **183**, 341.

White, M., Scott, D. and Silk, J. 1994, to appear in ARA&A.

White, S. D. M., & Rees, M. J. 1986, *MNRAS*, **183**, 341.

Woody, D. P. & Richards, P. 1981, *ApJ*, **248**, 18.

Zel'dovich, Y., & Sunyaev, R. 1969, *Ap. Space Sci.*, **4**, 301.