The Evolution of Density Fluctuations through Reheating in the Inflationary Universe and Gravitational Instability in the Non-linear Regime

by

Lam Hui

Submitted to the Department of Physics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

August 1996

Author ....
Department of Physics
August 20, 1996

Certified by ....
Alan H. Guth
Professor of Physics
Thesis Supervisor

Certified by ....
Edmund Bertschinger
Professor of Physics
Thesis Supervisor

Accepted by ....
George Koster
Chairman, Graduate Thesis Committee

The Evolution of Density Fluctuations through Reheating in the Inflationary Universe and Gravitational Instability in the Non-linear Regime

by

Lam Hui

Submitted to the Department of Physics
on August 20, 1996, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

ABSTRACT

Three separate problems are addressed. First, the effects of reheating on density fluctuations are examined in the framework of the inflationary cosmology. It is commonly believed that appropriately chosen gauge invariant measures of density fluctuations remain constant as long as the length scales of fluctuations greatly exceed the Hubble radius. Since scales of cosmological interest are much larger than the Hubble radius at the time of reheating, the behavior of such gauge invariant measures is often taken to imply that details of the reheating process do not affect density fluctuations on large scales. We demonstrate an exception using a simple instantaneous reheating model.

Second, we examine a class of approximations, referred to as local approximations, which can be applied in the nonlinear regime before trajectory-crossing. It is shown that the celebrated Zel'dovich approximation (Zel'dovich 1970) can be viewed as one in a series of approximations that involve the truncation of a hierarchy of Lagrangian fluid equations. They share the common property that the evolution of quantities such as density at each mass element is described by a closed set of differential equations decoupled from those of other mass elements. An approximation of this nature, which we call the Local Tidal Approximation (LTA), is introduced and is shown to be exact in cases of planar, cylindrical and spherical symmetries. The two approximations, along with a third example, are tested in the case of ellipsoidal collapse and the LTA is found to be a good approximation.

Third, the Zel'dovich approximation, with appropriate smoothing and coupled with the thermal and ionization evolution equations, is used to study the Lyman-α forest. It is assumed that the Lyman-α forest is a manifestation of density and velocity fluctuations which result from gravitational instability in hierarchical clustering cosmological models. The part of the forest that arises from mildly nonlinear regions can therefore be treated using a local approximation such as the Zel'dovich approximation. We introduce a method to compute the column density distribution.

Its dependence on the properties of the intergalactic medium as well as the power spectrum is investigated.

Thesis Supervisor: Alan H. Guth
Title: Professor of Physics

Thesis Supervisor: Edmund Bertschinger
Title: Professor of Physics

ACKNOWLEDGMENTS

獻給爸爸，媽媽。

First and foremost, I would like to thank my parents, who have always been encouraging, supportive and understanding. It still amazes me that they would happily support my going almost half-way around the globe to study something so incredibly remote from their lives, simply because they know this is what I like. I dedicate this thesis to *pa pa* and *ma ma*.

I thank my sisters La, Nam, Yan and my brother Chun-Kit, for being patient with and very supportive to their somewhat irresponsible brother. It was very fortunate indeed that Chun-Kit and La could come to visit me here in Cambridge in the course of my study and I still remember with fondness the time we have spent together. To Nam and Yan: I hope you can both come visit in the near future too.

I am very grateful to my thesis supervisors Ed Bertschinger and Alan Guth. It is my great fortune to be able to work with, not one, but two of the best scientists in the field. To be able to see their great minds at work, each in their own very special ways, is certainly the most valuable educational experience in my graduate career. Their encouragement to seek interesting problems and strike out my own path has been extremely important. At a time when funding for scientific research is so tight, their unfailing support for me to attend international conferences and schools is truly exceptional. These international experiences reinforce my desire to pursue the study of the universe, to join that community of truth-seekers. I also thank them for numerous comments and suggestions on the thesis. The work described in chapters 1 and 2 is done in collaboration with Alan and that in chapters 3 and 4 is done with Ed, although they have both given invaluable advice for all the chapters.

I am grateful to Fred Rasio for being the reader of the thesis and giving me useful comments. It is a pleasure to thank Nick Gnedin, my collaborator for the work in chapter 6. The thesis probably would not have been completed in time without his help. In the course of my research, I have benefited from discussions with many

scientists, among them: Robert Brandenberger, Shaun Cole, Leonid Grishchuk, Lars Hernquist, Bhuvnesh Jain, Sabino Matarrese, Hojun Mo, Roman Scoccimaro, Peter Schneider, Paul Steinhardt and Simon White. I am also indebted to Yu Zhang for kindly providing the results of his simulation, which are very important for testing the approximations I make in chapter 6, and Esther Hu for sending me her observational data on the column density distribution.

I would like to express my deep gratitude also to my undergraduate thesis supervisor Edgar Knobloch, whose faith in my ability encouraged me to pursue graduate study.

The thesis defense about two weeks ago is still fresh in my memory. I remember how overwhelmed I was to suddenly realize all the things my friends have prepared for me, from the preparation, defense to the celebration afterwards. It is fair to say that the thesis should not be viewed as the work of an individual but a collective effort of colleagues as well as friends.

I do not know how I can possibly thank my friend, office-mate and house-mate, Bob Rutledge enough. Through all the rough times and trials in my time at MIT, he has always been there. More than anyone else, he is the one who taught me how to program, to send e-mail, to use netscape ... introduce me to Once Blue, Trainspotting ... I thank my office-mates Dave, Derek, Eric, Jeff, Patrick, Tao-Tao for giving me help on Latex, taking messages and all the things office-mates do for each other. I am extremely grateful to Walter Lewin for allowing me to use the computer facilities in 37-616 and for giving me his support when I was seeking a postdoctoral position.

I am grateful to the building 6 folks Anand, John, Jim, Linquing, Marie, Matias, Paul Bode, Paul Schecter, Rennan and Uros for companionship and many useful discussions. I am especially indebted to Jim who has been my private (unpaid) computer counselor. His programming skills are truly amazing. I also remember he is the one who tried to print out a figure for me in, literally, the last minute before the thesis defense. I thank Charlie, Bill, Joe, Josh and Raissa for the time we spent together, in the cage and outside.

It would be wrong for me not to mention two secretaries who have done such a

superb job in helping us take care of various paper-work and giving us tips on things big and small. Thank you Chris. Thank you Peggy.

There are so many friends I have to thank, whose companionship, in addition to keeping me sane at the most insane times, also remind me there are things more important than my work. To Albert, Florence, Jackie, Kanas and Romy I thank for the dinners that we had together, the tennis games and sailing and camping trips. David, Choi, Edward (siulong), Joe, Jinha, Karen, Shuchung, Ronald, Yung-ming (Tai-lo): thank you for the many late-night conversations we had at one of our homes. My philosophy professor at Berkeley Paul Feyerabend once told me life comes in pockets of senses. Those are the pockets that make sense. Precious.

Then there are friends whose companionship is not immediate, because they are far away, but nonetheless very real. "No" thanks to Helen, as she always says. The friendships of Jin, Linda (Hess), Linda (Lai) and Micah are close to my heart. The old LeConte Physics gang, Lokesh, Matt, Morgan, Peter, Ramesh, Sim: we have not been in touch a whole lot but the visit of Matt a few weeks ago brought back a lot of fond memories. I thank my colleagues at NDC, Edith and Paul, Poon and Linda (Lee) for checking me out from time to time. To Ingrid, Hilda, Scott, Onion: our deep connection, despite the distance and infrequent communication, is both a source of strength and hope. I often have the faces of you in my mind as I travel to all the conferences on different continents.

Lastly, Shuk-shan, I know it is a word one should not use lightly, but thank you, thank you for your love.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

It is perhaps unusual that three separate problems of a different nature are brought together in a single thesis. One of them concerns the state of the universe at its first 10^{-35} second, when it was 10^{-27} times smaller that it is today and the temperature was around 10^{27} K or 10^{14} GeV. The universe was filled with a sea of relativistic particles and radiation pressure was significant. This is the beginning of what is commonly called the radiation-dominated era. The universe was very homogeneous then. The other two problems considered in this thesis concern the evolution of the universe from about 10^3 years (when its temperature was about 10^4 K or 1 eV) up to today. This is the era of matter-domination. Radiation pressure has become negligible. This is the period when the structures including, quasar absorption systems, galaxies, clusters of galaxies and superclusters formed out of the primordial cosmological fluid.

The common theme underlying these three problems is the evolution of density fluctuations. Today most cosmologists work within the paradigm of structure formation by gravitational instability. It is believed that the early universe was very homogeneous, except for tiny fluctuations, as is supported by measurements of the microwave background anisotropy (Smoot et al. 1991; Bennett et al. 1994). The amplitude of fluctuations was one part in 10^5 on large scales at the time when radiation decoupled from baryons. Structures that we see today form from the growth of the density fluctuations under gravity. Overdense regions collapse to form galaxies and clusters. Matter moves away underdense regions which turn into voids. This frame-

work is adopted in the present thesis. Alternative theories are discussed in Peebles (1993).

There are then two questions the cosmologist must address. The first is the specification of initial conditions. What was the spectrum of the primordial fluctuations? How to explain its observed amplitude? Answers to these questions require knowledge of high energy physics which is not tested in terrestrial experiment. Plausible theories exist, which are largely based on the idea of phase transitions in the early universe. Combining them with the gravitational instability paradigm, predictions for structure formation are made which can be tested against observations. The theory of inflation is discussed in the first part of the thesis. In particular we focus on possible effects of the phase transition between the inflationary and the radiation-dominated stages on the density fluctuations.

The second question is one of evolution.[1] Given the initial conditions predicted by plausible theories, the evolution of density fluctuations can be divided roughly into three different stages.[2] The first stage is when the fluctuations evolve according to linear theory because they have very small amplitudes. This holds, for fluctuations on scales of cosmological interest, from the the end of inflation, through the radiation-dominated era and recombination to some point in the matter-dominated era. From this time (which depends on the scales one is interested in) onward, the fluctuations continue to grow under gravitational instability until gas pressure and/or virialization becomes important. The evolution of fluctuations during the period when the overdensity (density fluctuation divided by mean density) is comparable to 1 but not too much bigger can be treated using approximate methods. In the second part of the thesis, we define a class of such approximations, referred to as local approximations. We show that the celebrated Zel'dovich approximation (Zel'dovich 1970) is one

[1]Strictly speaking, the first problem of initial conditions also involve evolution itself: how do density fluctuations evolve through the phase transition that ends inflation? One can view the initial condition problem as the problem of specifying the fluctuation at Hubble-radius crossing after inflation.

[2]It should be emphasized that the stages we refer to are scale dependent. At any given time, there are scales on which the fluctuations are in any one of the stages. In hierarchical clustering cosmological models, the smaller scales tend to be in the more nonlinear stages at a fixed time.

example of it. Another example is discussed and a new local approximation is proposed and tested. The local approximations can be applied in the mildly nonlinear regime before orbit-crossing. The third stage is one in which nonlinear effects are so important that the approximate methods fail. In the strongly nonlinear regime, N-body simulations provide perhaps the only reliable means to compute the evolution of density fluctuations of cold (i.e. pressureless) matter.[3] Hydrodynamics also has to be incorporated to take into account the effects of gas pressure and shocks. These powerful computational methods are not discussed in this thesis, although use is made of some of the results obtained using them.

In the third part of the thesis, the Zel'dovich approximation is applied to study the part of the Lyman-α forest that is in the mildly nonlinear regime. We develop the methodology to use the approximation to compute, for any given cosmological model, the distribution of neutral hydrogen column density, which is a commonly used statistical measure of the Lyman-α forest. It is demonstrated how this can be used to constrain cosmological models, which brings us a full circle back to the first problem of initial conditions.

The following discussion is accordingly divided into three different parts, on the subjects of inflation, local approximations and the Lyman-α forest respectively. We state the questions to be addressed in the subsequent chapters and give a brief introduction to the basic ideas of each field. Some background material on the Friedmann-Robertson-Walker cosmology (Friedmann 1922; Robertson 1935; Walker 1936) is given in the section on inflation. For general references on cosmology and structure formation, the reader is referred to several excellent monographs (Weinberg 1972; Zel'dovich & Novikov 1983; Peebles 1993; Padmanabhan 1993).

[3] A possible exception is the application of the Press-Schecter theory (Press & Schechter 1974) to study the statistics of cold matter halos.

1.1 The Inflationary Universe

Inflation was first formulated in a clear manner by Guth (1981) who emphasized how a brief period of almost exponential expansion in the early universe, motivated by ideas of Grand Unification, could solve a host of fundamental cosmological problems. The original model did not work, as realized by Guth himself. Alternative formulations avoiding the problem in the original model were proposed by by Linde (1982) and Albrecht and Steinhardt (1982). Since then, numerous models, each realizing the original ideas in a different way, have been proposed. Unfortunately, we do not understand high energy physics (around 10^{14} GeV) well enough to pin down exactly which model is the correct one. However, they all share certain common features, on which we concentrate.

Shortly after the first successful inflationary models were proposed, it was realized by several groups that besides solving the cosmological problems presented by Guth (1981), all of which, one way or another, had to do with why the universe was so smooth at an early time, inflation could also provide an explanation of why there had to be small departure from smoothness (Starobinsky 1982; Hawking 1982; Guth & Pi 1982; Bardeen et al. 1983). There are roughly speaking two parts to the calculation of how inflation gives rise to density fluctuations in the later stages of the universe, one making use of quantum physics and the other making use of general relativistic perturbation theory. It suffices to say that the root-mean-squared amplitude of a fluctuation on a given scale at the earliest moments is dictated by quantum mechanics while the subsequent evolution through the end of inflation to the radiation-dominated and matter-dominated stages is governed by general relativity. Subtle questions arise concerning the quantum-to-classical-transition which is a whole subject in its own right The reader is referred to Guth and Pi (1985) and references therein.

We focus on the general relativistic part of the calculation. In particular, we examine how reheating or the transition between the inflationary and radiation-dominated stages can affect the density fluctuations. It is commonly believed that

one can define certain measures of fluctuations which determine the density fluctuations in the post-inflation-stages and which remain constant throughout reheating on scales of cosmological interest (i.e. much larger than the Hubble radius at the time of reheating). The implication is that reheating has no physical effect on the density fluctuations we can observe today, the reasoning being that physically significant quantities cannot change on scales larger than the Hubble radius. This view is challenged in Chapter 2 and 3 of the thesis. A hypothetical situation is presented in which inflation ended suddenly at some point when all fluctuations of cosmological interest have wavelengths much larger than the Hubble radius. It is shown that the amplitude of fluctuations in the later stages can be altered significantly depending on the manner in which inflation is ended. The work was motivated by a claim of Grishchuk (1994) that the standard calculation regarding the evolution of density fluctuations is flawed. It turned out that an error was made by him in the junction conditions used in connecting the fluctuations across the transition between the inflationary and radiation-dominated stages, which has been pointed out by Deruelle and Mukhanov (1995). While working independently on the same subject, we discovered a different problem of the standard calculation which forms the subject of the first part of this thesis.

We give below a quick tour of the major concepts that lead to the subject of investigation. Two excellent reviews on the inflationary cosmology are recommended: Blau and Guth (1987) and Linde (1990). For a comprehensive review on the subject of density fluctuations in the inflationary universe, see Mukhanov, Feldman and Brandenberger (1992).

1.1.1 The Horizon Problem

The standard hot big bang cosmology [4] is known to have several flaws when it is extrapolated to early times. One of them is the horizon problem. There are two basic equations that describe the evolution of the scale factor of the universe (in other

[4]By standard hot big bang cosmology we mean Friedmann-Robertson-Walker cosmology without inflation.

words, its size) in standard Friedmann-Robertson-Walker cosmology:

$$\frac{1}{2}\left(\frac{da}{dt}\right)^2 - \frac{G}{a}\left(\frac{4\pi}{3}\rho_0 a^3\right) = -\frac{C}{2}, \tag{1.1}$$

$$\frac{d\rho_0 a^3}{dt} = -p_0 \frac{da^3}{dt}, \tag{1.2}$$

where a is the scale factor, ρ_0 is the average density and p_0 is the average pressure of the universe. t is proper time and C is a constant. G is Newton's constant.

Both equations can be viewed as expressions of energy conservation. In the first equation, the first term on the left hand side is the kinetic energy of a particle of unit mass moving with velocity $\dot{a}$ and and the second term is the potential energy associated with such a particle if it is situated at the surface of a sphere of radius a. $-C/2$ is then simply its total energy, which remains constant. (C is commonly known as the curvature.) The second equation resembles statements of the first law of thermodynamics: change in energy ($a^3 d\rho_0$) equals minus pressure ($-p_0$) times the change in volume (da^3).

Given any equation of state that relates density and pressure, one can solve the above two equations for the evolution of a. In the universe today, pressure is negligible on large scales (matter-dominated). It then follows from equation (1.2) that $\rho_0 \propto a^{-3}$ which when substituted into equation (1.1) and assuming $C = 0$ (which corresponds to a flat universe, a point we will return to in a minute), implies that $a \propto t^{2/3}$. At earlier times, when the universe was hot and radiation pressure was important, $p_0 = \rho_0/3$. It can be shown similarly that $a \propto t^{1/2}$. [5]

The horizon is defined as the proper distance a photon can travel from the hot

[5]Notice that we have not been entirely rigorous by saying that $a \propto t^{2/3}$ in the matter-dominated era. Suppose $t = 0$ is the moment of the hot big bang. Then during the radiation-dominated era $a \propto t^{1/2}$ holds. But in the matter-dominated stage, a should be equal to $const.A(t - const.B)^{2/3}$ where the const.A and const.B are fixed by matching a and $\dot{a}$ at the transition between the two era. However, for the purpose of the discussion here, it suffices to think of $a \propto t^{2/3}$ in the matter-dominated stage. In any case, for large enough t, this is a good approximation.

big bang up to some time of interest (t):

$$\text{Horizon} = a(t) \int_0^t \frac{c\, d\bar{t}}{a(\bar{t})}. \tag{1.3}$$

$c\, d\bar{t}$ is the proper distance the photon travels in time interval $d\bar{t}$; dividing it by a gives the comoving distance traversed in the same time interval; summing over all the little time intervals and multiplying by $a(t)$ gives the total proper distance covered measured at time t. For $a \propto t^{2/3}$ (matter-dominated), the above integral is proportional to $a^{3/2}$ and for $a \propto t^{1/2}$ (radiation-dominated), it is proportional to a^2. The important feature to bear in mind is that the horizon in both cases grows faster than a.

Imagine a fluctuation today of a given wavelength. Its proper wavelength grows like a (i.e. it grows more slowly than the horizon does).

This is the horizon problem: for a fluctuation of wavelength smaller than the horizon today, at some sufficiently early time, the same fluctuation had a wavelength which exceeded the horizon at that time. How, then, did it form? [6]

In fact, it can be shown that at sufficiently early times, the required amplitude of fluctuations on scales larger than the horizon was much smaller (but non-zero) than what one would expect from pure Poisson fluctuations (unless the number of particles becomes arbitrarily large in the early universe). No known causal process can possibly explain the origin of such fluctuations in the framework of standard hot big bang cosmology. This does not mean such fluctuations violate known laws of physics. It just mean that we cannot explain them except to postulate them as initial conditions given by some process beyond the standard model - or by the Creator.

Before we leave this section, let us quantify the horizon problem. In the matter-dominated era, as we have discussed, the horizon grows like $a^{3/2}$ or t. The reciprocal of the Hubble parameter (commonly known as the Hubble length or Hubble

[6]The horizon problem is often formulated as the riddle of why the universe was so smooth on scales much larger than the horizon at some early time. We formulate it in terms of fluctuation instead of smoothness but of course they are two sides of the same coin. The problem is why there existed fluctuations on large scales on the one hand and why they were so small on the other.

radius) , $(\dot{a}/a)^{-1}$, has the same time dependence and can be taken to be a measure of the horizon. The same is true of the radiation-dominated era. Today the Hubble parameter is of the order of $100\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ which translates into a size of $10^{28}\,\mathrm{cm}$ (cH^{-1}). At $10^{14}\,\mathrm{GeV}$ the universe had a temperature around $10^{27}\,\mathrm{K}$, which was about 10^{27} times hotter than it is today.[7] This means the ratio of the scale factor a today to a at $10^{14}\,\mathrm{GeV}$ is also 10^{27}. Assuming the Hubble parameter has been growing like $a^{3/2}$ all the way back to $10^{14}\,\mathrm{GeV}$ (which is of course not true because by then the universe was already radiation-dominated and the growth rate was a^2 but for our present argument, making a small error in the order of magnitude estimate is immaterial), this implies that cH^{-1} was of the order of $10^{-12}\,\mathrm{cm}$ at that time. On the other hand, galactic scale today is around $10^{22}\,\mathrm{cm}$ and so at $10^{14}\,\mathrm{GeV}$, the same fluctuation had proper length of $10^{-5}\,\mathrm{cm}$, [8] a full 7 orders of magnitude larger than the Hubble radius! Considering larger scale fluctuations today such as clusters, voids or even the cosmic microwave anisotropy only exacerbates the problem.

1.1.2 The Inflaton

The canonical inflationary model postulates that at some early stage, the energy density of the universe was dominated by that of a scalar field, which we call the inflaton. The mean energy density and pressure of the inflaton had the following forms:

$$\rho_0 = \frac{1}{2}\dot{\varphi}_0^2 + V(\varphi_0) \tag{1.4}$$

$$p_0 = \frac{1}{2}\dot{\varphi}_0^2 - V(\varphi_0) \tag{1.5}$$

[7] We choose $10^{14}\,\mathrm{GeV}$ because this is the canonical time when inflation was supposed to end. The actual time when inflation ended can vary somewhat depending on the model. At earlier times, the horizon problem could only be worse.

[8] We assume the comoving length remains constant and the proper length simply grows like a, which is almost certainly not true for galactic scale objects. One can think about larger scale fluctuations for which the assumption holds, but they are bound to create a worse horizon problem i.e. they were even longer compared to the horizon at $10^4\mathrm{GeV}$.

where φ_0 is the mean scalar field and V is its potential. The dot denotes differentiation with respect to proper time t. Suppose the potential was of the form shown in Fig. 1-1. Let us further assume that the scalar field was initially very close to the origin and had very little kinetic energy at first. Then according to the above equations, $\rho_0 \sim V(0)$ and $p_0 \sim -V(0)$. Eventually the scalar field rolled away from the origin towards the true minimum of the potential, gaining kinetic energy along the way. But there was a period of time in which the energy density was to first approximation a constant and the pressure was negative.

The implication is enormous if at some stage in the early universe, the above conditions were satisfied. Recall equations (1.1) and (1.2). The latter is automatically satisfied when energy density and pressure are both constant but have opposite signs. The former implies that

$$a = \frac{\sqrt{-C}}{H_{\text{infl}}} \sinh H_{\text{infl}}(t - t_i) \qquad (\text{for } C < 0)$$
$$a = \frac{\sqrt{C}}{H_{\text{infl}}} \cosh H_{\text{infl}}(t - t_i) \qquad (\text{for } C > 0) \tag{1.6}$$

where H_{infl} is defined by $H_{\text{infl}}^2 = 8\pi G V(0)/3$ and t_i is simply a constant, which can be regarded as the beginning of inflation (we will take inflation to mean the whole period during which the scalar field dominated the energy of the universe). For sufficiently large $H_{\text{infl}}(t - t_i)$ (i.e. if the inflaton was perched at the top of the potential for a long enough time), a was to excellent approximation proportional to $\exp H_{\text{infl}}(t - t_i)$, for both solutions.

The early evolution of a was then very different from that of the standard hot big bang theory. Instead of power law growth (t^2) as in the radiation-dominated era, we had exponential growth, hence the name inflation.

It should be emphasized though that because the inflaton did move, albeit slowly, away from the top of the potential, the energy density was not exactly constant (equation [1.4]). Hence the expansion was not exactly exponential. But during the first stage when the inflaton had very little kinetic energy, the expansion was well approximated by exponential. This period of inflation is called the slow-roll stage.

The magic of the exponential could be exploited to yield huge amount of expansion, even with relatively modest amount of time passage ($H_{\rm infl}(t - t_i)$ of the order of 50 say).

An immediate consequence is that the ratio of $C/2$ to the potential energy $4\pi GV(0)a^2/3$ in equation (1.1) became negligible by the end of inflation. That is to say, the kinetic and potential energy of expansion almost balanced each other. This is known as flatness. In fact it can be shown that without fine tuning parameters of the potential, the amount of inflation was sufficient to keep the universe flat even up to today, to excellent approximation. This is the famous generic prediction of flatness.

It is easy to get carried away when one discusses all the goodies of inflation. We will turn our attention to the solution of the horizon problem offered by inflation, because this leads us to the main concern of this thesis: generation and evolution of density fluctuations. Before we do so, however, let us pause and discuss the motivation for the domination of the inflaton in the early universe.

In any scheme of grand unification, the underlying gauge field theory always had certain symmetries at early times when the temperature of the universe was high enough but they were broken later on as the universe cooled. An example of such symmetry breaking, which has come to be adopted as part of the standard model, is the separation of the electroweak interaction into electromagnetism (where the gauge boson, the photon, is massless) and the weak interaction (where the gauge bosons, the W's and the Z, are massive). The idea of symmetry breaking is quite simple and deep. Recall that in Fig. 1-1, the potential is symmetric about zero. Suppose that at high temperature, the potential was still symmetric about zero but had a different form such that $\varphi_0 = 0$ was actually the true minimum. The scalar field then naturally sat at the origin. At lower temperature, the potential changed shape to one that resembled that in Fig. 1-1. The origin became an unstable minimum. Quantum fluctuations alone could cause the scalar field to roll away from it and eventually settle at the true minimum, which was at $\varphi_0 \neq 0$, thereby breaking the symmetry. In electroweak theory, the symmetry is not the same as the one in Fig. 1-1 but the idea is quite similar. Suppose further that the scalar field was actually coupled to

other fields, for instance, the W's and the Z. The fact that the scalar field took an expectation value away from the point of symmetry caused these particles to gain masses. In the case of the electroweak theory, the agent of symmetry breaking (the scalar field) is the Higgs field. Other attractive features of the idea of symmetry breaking include renormalizability, which we will not discuss here.

The inflaton can be viewed as the analog of the Higgs for unification at higher energy scale. The canonical energy scale at which symmetry breaking occurs is around 10^{14}GeV. Unfortunately, there are a lot of grand unified theories and we do not know which is correct, if any. However, it is not unreasonable to expect that the universe underwent phase transitions such as the one described above. If the energy associated with the agent of symmetry breaking dominated that of the universe at some point, something like inflation had to occur.

Eventually, this period of inflation came to an end. Otherwise, we would have been left with a very cold universe due to the enormous expansion. This happened after the inflaton picked up kinetic energy and rolled into the valley of the potential. As it oscillated around the minimum of the potential, it lost energy through coupling to other particles. The universe reheated. By the time the inflaton lost most of its energy and settled into the true minimum of the potential, the radiation-dominated era of the standard hot big bang theory began. The period of inflaton decay is called reheating.

1.1.3 Solution to the Horizon Problem

Recall that we have shown the Hubble radius at 10^{14} GeV was around 10^{-12} cm while galactic scale fluctuation had a length scale of around 10^{-5} cm at the same time. The horizon problem is solved because inflation implies that the Hubble length of 10^{-12} cm was really not the horizon size at 10^{14} GeV.

Suppose that at the beginning of inflation, the Hubble radius was also around 10^{-12} cm (remember that during inflation, because of the almost constant energy density, the Hubble parameter was also almost constant). Suppose further that 10^{-12} cm was indeed the horizon size *at the beginning of inflation.* Then the galactic scale

fluctuation when extrapolated to the start of inflation, (i.e. shrink its proper length 10^{-5} cm further by the amount of expansion during inflation) could have a length scale smaller than the horizon size at that time with, say, only 17 e-foldings of expansion. Double the number of e-foldings and the entire observable universe today was within the horizon at the beginning of inflation! With 50 e-foldings, the observable universe plus much more was within the horizon. Such is the power of the exponential.

Note that we can repeat the same argument without assuming the Hubble radius corresponded to the horizon size at the beginning of inflation at all. After all, we really do not know very much about the state of the universe at such high energy scales. But given any horizon size at the beginning, a sufficient amount of inflation can always be chosen so that the observable universe was within the horizon then.

Notice that by the end of inflation, the Hubble radius, 10^{-12} cm, was not the horizon size anymore. To see that, suppose again that the horizon at the *beginning of the inflation* had a proper length of 10^{-12} cm, a lower bound on the horizon size at the *end of inflation* is simply 10^{-12} cm multiplied by the amount of expansion during inflation. This is because the horizon at any given time is equal to the Hubble scale factor at that time multiplied by the comoving distance a photon has traversed by then. The comoving distance a photon had traveled by the end of inflation was certainly larger than it was at the beginning. This distance (the comoving length that the photon has traversed before inflation began), when multiplied by the Hubble scale factor at the end of inflation, was precisely 10^{-12} cm multiplied by the amount of expansion during inflation. This provides a lower bound to the horizon size at the end of inflation. The horizon was therefore much larger than the Hubble radius after inflation.

Within the framework of inflationary cosmology, fluctuations on all observable scales today were actually within the horizon at early times. This means that it is possible to explain them using known causal processes.

1.1.4 Density Fluctuations: the Standard Lore and its Problem

Fig. 1-2 depicts the evolution of a fluctuation in the inflationary cosmology. What inflation achieves is to destroy the equality of the horizon size and the Hubble radius, thereby allowing the wavelength of the fluctuation to be much larger than the Hubble radius but smaller than the horizon at the end of inflation (which can be taken to be the beginning of the standard hot big bang universe).

Consider a fluctuation which still had proper wavelength much smaller than the Hubble radius after some amount of inflation had already taken place. The universe was very cold. The inflaton was settled at its ground state. By this we do not mean the inflaton had already settled at the true minimum of the potential. On the contrary the inflaton was perched at the top of the potential. If one decomposes the inflaton field into Fourier modes, each mode behaved like a simple harmonic oscillator. Each was in its ground state because the amount of energy required to excite it was much larger than the thermal energy available.

Quantum mechanics tells us that even in its ground state, the inflaton had inevitable fluctuations due to the uncertainty principle. This led to energy or mass fluctuations because the inflaton itself carried energy. Furthermore, because of coupling to gravity, metric perturbations arose as well.

Because of inflation, the proper wavelengths of these fluctuations grew by enormous amounts. Eventually, they were stretched beyond the Hubble radius. By the end of inflation we thus had a whole set of fluctuations (those on larger scales having exited the Hubble radius at earlier times), the wavelengths of which were much larger than the Hubble radius, but of course, smaller than the horizon.

In this thesis we will concentrate on the evolution of the fluctuations from the moment they left the Hubble radius until the end of inflation (see Fig. 1-2).

The proper framework for discussing the evolution of density fluctuations in the early universe is provided by general relativity. Since the amplitude of the primordial fluctuations was small, linear perturbation theory (i.e. throwing away terms that are

second or higher order in density, pressure or metric perturbations) is sufficient for our purpose. By virtue of general covariance, one is free to pick any space-time coordinates to follow the perturbations. This is commonly known as gauge freedom. Different observers can see different things. For instance, density fluctuations can be zero in one gauge but not in another. Alternatively one can choose certain combinations of the perturbed quantities which are gauge invariant to first order. One useful example, to which we will return again and again is

$$\zeta \propto a^2 \left(H_{\rm local}^2 - \frac{\rho}{3} \right) \tag{1.7}$$

where ρ is the full energy density (mean density plus its fluctuation), a has its usual meaning as the Hubble scale factor and $H_{\rm local}$ is a measure of the local expansion rate. The definition of ζ here agrees with that by Mukhanov, Feldman and Brandenberger (1992). The proportionality factor left out has no time dependence. These quantities are defined more carefully in Chapter 2. What is important here is the resemblance of the above with equation (1.1). If one equates the H_{local} above with the familiar Hubble parameter ($\dot{a}/a$), comparing with equation (1.1), it is not hard to see that ζ is then simply a constant proportional to the curvature C.

Now, equation (1.7) is, of course, not equivalent to equation (1.1). After all the former is meant to describe fluctuations about the mean while the latter refers to the mean itself. ρ in equation (1.7) contains fluctuations while ρ_0 in equation (1.1) the other does not (it is the average value). Similarly, while $\dot{a}/a$ describes the global expansion rate, H_{local} is defined as the local divergence of velocities, which can vary from one place to another. Therefore, ζ is a measure of the local curvature (as opposed to global curvature C). Put it in another way, it quantifies by how much a given local region deviates from flatness.[9]

ζ is a useful quantity because a particularly simple equation can be written

[9]It turns out, however, that ζ really carries two modes, one of them is like the local curvature as suggested. The other is the subject of almost half the thesis.

down for its evolution during inflation (its proof is deferred to Chapter 2):

$$(a\sqrt{\gamma}\zeta)'' + \left(k^2 - \frac{(a\sqrt{\gamma})''}{a\sqrt{\gamma}}\right)(a\sqrt{\gamma}\zeta) = 0 \tag{1.8}$$

where the prime (as opposed to dot) denotes derivative with respect to conformal time τ (conformal time and proper time are related by $ad\tau = dt$), k is the wavenumber of the particular fluctuation of interest and γ is defined by

$$\gamma = \frac{\varphi_0'^2}{2a^2H^2} \tag{1.9}$$

where φ_0, to remind the reader, is the mean inflaton field and H is the global Hubble parameter $\dot{a}/a$.

When a fluctuation was within the Hubble radius during the early phase of inflation, k^2 was large (small wavelength limit) compared to $(a\sqrt{\gamma})''/(a\sqrt{\gamma})$. (This assumes something about how γ behaves. We will discuss this in detail in Chapters 2 and 3) Equation (1.8) can be approximated by $(a\sqrt{\gamma}\zeta)'' + k^2 a\sqrt{\gamma}\zeta$, which is the equation of a simple harmonic oscillator of frequency k^2. It fits in nicely with our idea that the perturbation originated as the fluctuation of a quantum mechanical simple harmonic oscillator in its early phase. This can be made more precise by noting that there is in fact an exact quantum analog of the same equation with ζ being an operator (Mukhanov et al. 1992).

Eventually, because of the enormous stretching by inflation, the same fluctuation left the Hubble radius. k^2 was then small compared to the other term $(a\sqrt{\gamma})''/(a\sqrt{\gamma})$ in equation (1.8). Ignoring the k^2 term, one obvious solution of the equation is ζ being a constant. If one can show that the other solution decayed (remember equation (1.8) is second order and so has two independent solutions), ζ would approach a constant.

This is in fact the standard lore. Any given fluctuation mode, embodied by the variable ζ, oscillated in its "quantum mechanical phase" when it was within the Hubble radius. After it exited the Hubble length, it stopped oscillating and the amplitude became eventually frozen. In fact, its amplitude remained unchanged

until the same fluctuation re-entered the Hubble radius at a later time after the end of inflation (see Fig. 1-2). Afterwards, the fluctuation could grow in the matter-dominated era, which gave rise to structures we see today.

The non-constant mode of equation (1.8) can be written as (assuming k^2 can be ignored) $\zeta \propto \int_{\tau_i}^{\tau} d\bar{\tau}/a^2\gamma$ where τ_i is a constant. Note how the expression becomes problematic if γ vanishes. According to equation (1.9), γ could be zero when the inflaton stopped moving.

φ_0' (or γ) was small but non-vanishing during the initial phase of inflation, when the inflaton moved slowly away from the top of the potential. However, after the inflaton rolled into the valley of the potential, it picked up speed, went past the minimum of the potential and eventually came to a stop. We will call it the point of turn-around from now on. At this particular moment, γ became strictly zero and it is not obvious how ζ behaved. In particular, did its amplitude remain frozen through the point of turn-around? Did it diverge at the moment of turn-around and what happened after it? What is the implication for density fluctuations today and under what the circumstances does the standard lore break down?

These questions will be addressed in Chapters 2 and 3.

1.2 Local Approximations

In the second part of the thesis, a set of approximations to the gravitational collapse of cold (pressureless) matter are discussed. They share a common property which we call locality. By this we mean that each mass element behaves as if it is independent of all the others. To be more precise, properties of each element, such as its density, velocity gradient and so on, are described by a set of ordinary differential equations which are decoupled from those describing other elements.

The present work was motivated by that of Bertschinger and Jain (1994) who examined the consequences of what we call the Non-magnetic Approximation (NMA). They used the fact that if a quantity known as the magnetic part of the Weyl tensor vanishes in the Newtonian limit, a set of exact Lagrangian fluid equations for cold

dust becomes local. This was first proven in general relativity by Matarrese, Pantano and Saez (1993) following earlier work of Barnes and Rowlingson (1994); part of the motivation for such an assumption was the statement of Ellis (1971) that there is no counterpart to the magnetic part of the Weyl tensor in Newtonian theory. One important result they obtained is that spindle (filamentary) collapse is favored in general as opposed to pancake collapse. (Pancake collapse had been thought — correctly — to be the generic outcome of gravitational collapse of cold dust following the work of Zel'dovich 1970.)

Since then, it has been shown by Bertschinger and Jain (1994) and Kofman and Pogosyan (1995) that the magnetic part of the Weyl tensor does have a Newtonian counterpart and so spindle collapse is not favored by the exact dynamics. Kofman and Pogosyan (1995) also showed that the tidal evolution equation implied by the Zel'dovich approximation is local.

In this thesis, we extend the above work by first showing that the celebrated Zel'dovich approximation can be formulated entirely in the language of local Lagrangian fluid equations. The aim is to make it easy to search for generalizations of the Zel'dovich approximation, which is known to be an excellent approximation in the mildly nonlinear regime before orbit-crossing. Are there other local approximations that incorporate one of the successful features of the Zel'dovich approximation, namely that it is exact in the case of one-dimensional collapse or planar symmetry, and perhaps extends the property of exactness to cases of cylindrical and spherical symmetry as well? Why does the NMA result in spindle collapse generically? Are there other local approximations, other than the Zel'dovich approximation, that do not share the same feature? These questions are addressed in Chapter 4 and 5. A new local approximation is put forward and is compared with the previous two approximations in the case of general ellipsoidal collapse.

Two important remarks are in order here. The first is that the local approximations can be applied only up to trajectory crossing. However, it is true that for hierarchical clustering cosmological models, the approximations can be still applied, with appropriate smoothing, on large enough scales which are in the mildly nonlinear

regime, even if trajectory crossing has already occurred on small scales. This procedure has been shown to work well in the context of the Zel'dovich approximation (Coles et al. 1993). Peebles (1993) gave a nice summary of the history of ideas evolving around the original proposal by Zel'dovich. An excellent review of the Zel'dovich approximation and the larger subject of structure formation can be found in the article by Shandarin and Zel'dovich (1989). Our second remark is that although some of the early ideas on local Lagrangian fluid equations originate from work on general relativity, the discussion of local approximations can be (and is) held entirely in the Newtonian framework. For the applications of local approximations to general relativistic problems, the reader is referred to work by Bruno, Matarrese and Pantano (1995) and Matarrese and Terranova (1995).

In the following, we illustrate the idea of local approximations with an extremely simple example.

1.2.1 Locality

The equations that describe a pressureless, self-gravitating fluid in an expanding universe are:

$$\frac{d\delta}{d\tau} + (1+\delta)\nabla_i v^i = 0 \ , \tag{1.10}$$

$$\frac{dv^i}{d\tau} = -\frac{a'}{a} v^i - \nabla^i \phi \ , \tag{1.11}$$

$$\nabla^2 \phi = 4\pi G a^2 \rho_0 \delta \ . \tag{1.12}$$

Mean mass density is denoted again by ρ_0. δ is the overdensity defined by $\delta = (\rho - \rho_0)/\rho_0$ where ρ is the full mass density. $\vec{v} = d\vec{x}/d\tau$ is the proper peculiar velocity where $\vec{x}$ is the comoving spatial position and τ is the conformal time. We neglect spatial curvature so that Cartesian coordinates can be used such that $\nabla^i = \nabla_i = \partial/\partial x^i$ for the ith spatial coordinate. ϕ is the gravitational potential. The prime denotes derivative to conformal time as before. The notation $d/d\tau$ stands for:

$$\frac{d}{d\tau} \equiv \frac{\partial}{\partial \tau} + v^i \nabla_i \tag{1.13}$$

This is called the Lagrangian derivative. It denotes the time derivative of a quantity following a fluid element, as opposed to at a fixed spatial position ($\partial/\partial\tau$).

The description here is Newtonian except for the expanding cosmological background. Equation (1.10) expresses mass conservation. Note how the rate of change of density depends on the local divergence of the velocity field. The second one is the equation of motion. It describes how a mass element moves under the influence of the background expansion (first term on the right hand side) and the gravitational attraction of other elements (second term on the right). Equation (1.12) is the Poisson equation that relates overdensity to the gravitational potential.

The terms that contain spatial gradients are what we shall call non-local ones. Their presence means that to evaluate the time evolution of quantities of interest at one mass element, it is necessary to have information of the neighboring ones as well.

If gravity were absent, it would be easy to solve this problem. For instance, suppose one is interested in the evolution of δ at one particular mass element. Equation (1.10) tells us that the local divergence of the velocity field has to be known. But one can write down the evolution equation for the local divergence of the velocity field by taking the gradient of the equation of motion:

$$\frac{d\nabla_i v_j}{d\tau} = -\nabla_i v^k \nabla_k v_j - \frac{a'}{a}\nabla_i v_j - \nabla_i \nabla_j \phi \ , \tag{1.14}$$

Summing over i and j gives us the evolution equation for the divergence of the velocity field. It can be seen immediately that the divergence of the velocity field is coupled to other components of the velocity gradient tensor $\nabla_i v_j$, the evolution of which is also described by the above equation. Note that if gravity is absent ($\nabla_i \nabla_j \phi = 0$), equation (1.14) together with the mass conservation equation (1.10) form a closed set of equations that can tell us how the overdensity δ and the velocity gradient tensor $\nabla_i v_j$ evolve at any given mass element without having to know what others are doing. This is true as long as one is given the initial conditions on δ and the velocity gradient tensor at the mass element of interest. It is our first example of locality.

Without gravity, the above is almost a silly exercise. After all, with no gravitational interaction (i.e. the last term on the right hand side of equation [1.11] is absent), each particle simply moves in a straight line without caring about others. Naturally, locality is observed.

With gravity however, one has to know the distribution of all mass elements in order to calculate the gravitational acceleration at any one of them. The Poisson equation has to be solved, which is the biggest obstacle to locality. In other words, one has to think about how to deal with the last term on the right hand side of equation (1.14).

There are a lot of approximations one can make to restore locality. Each of them in one way or another eliminates the need of solving the Poisson equation. With locality, one can deal with one mass element at a time, which greatly simplifies the computation of predictions for large scale structure. It is perhaps surprising any local approximation could approach reality at all.

In Chapters 4 and 5, we show two known examples and propose a new local approximation. That one of them, the Zel'dovich approximation, is very successful, gives us the motivation to look for even better approximations along the same lines.

1.3 The Lyman-Alpha Forest

The third part of the thesis is on the Lyman-α forest, a whole separate subject in its own right. We do however make use of the Zel'dovich approximation described in the second part of the thesis. It is also argued that observations of the Lyman-α forest has some bearing on the initial conditions in the early universe.

The Lyman-α forest, which refers to a subset of the numerous Lyman-α absorption lines in quasar spectra, was first discovered by Lynds (1971). The main properties of the forest are summarized as follows. First, the Lyman-α forest consists of those Lyman-α absorption lines with column densities of neutral hydrogen less than about $10^{17}\mathrm{cm}^{-2}$. Above this column density, the optical depth of the Lyman-α absorption greatly exceeds one. The Lyman-limit systems and the damped Lyman-α

systems belong to this later category, and they usually have associated metal line absorption. Second, the comoving number density of the forest absorption lines increases with redshift (Peterson 1978; Lu et al. 1991b). Third, no significant spatial correlation beyond that of a random distribution has been detected between the forest absorption lines at velocity separations larger than about $250\,\mathrm{km\,s^{-1}}$ [10], implying they are quite uniformly distributed (Sargent et al. 1980). Based on this observation, it is commonly believed that the Lyman-α forest is not associated with galaxies. Weak clustering has been reported at lower velocity separations, a finding that was made possible by high resolution studies of C IV systems that were known to be associated with the Lyman-α forest at column densities higher than about $10^{14.5}\,\mathrm{cm^{-2}}$ (Songaila & Cowie 1996; Fernandez-de Soto et al. 1996). Fourth, the number of absorption lines per unit column density per unit redshift is approximately proportional to $N_{HI}^{-\beta}$ where N_{HI} is the column density and $1.4 \lesssim \beta \lesssim 1.6$ although there exists evidence of at least one break in the power law somewhere in the range $10^{14} - 10^{17}\,\mathrm{cm^{-2}}$ (Petitjean et al. 1993; Hu et al. 1995; Lu et al. 1991a). Fifth, the number of absorption lines per unit equivalent width (equivalent width W is the integral of the transmission over the wavelength range of an absorption line and has the unit of $\mathring{A}$ usually) per unit redshift $\propto \exp -W/0.3\mathring{A}$ except for a possible excess at equivalent widths less than $0.2\mathring{A}$ (Murdoch et al. 1986). It is commonly assumed that the above approximate relations hold independent of redshift. Sixth, the b-parameter (or b-value) , which is a measure of the width of a line, is observed to have a wide spread at a fixed column density but there is a lower limit at about $15 - 20\,\mathrm{km\,s^{-1}}$ (Songaila & Cowie 1996; Lu et al. 1991a) at $z \sim 3$. The precise definitions of column density, equivalent width and b-value are given in Chapter 6.

Obviously we cannot list all the interesting properties of the Lyman-α forest here. The reader is referred to a review article by Weymann (1993) for more information.

There are of the order of a hundred Lyman-α absorption lines in each quasar absorption spectra. Multiplying that by the number of quasars that can be observed,

[10] velocity separation is equal to the speed of light times $\Delta\lambda/\lambda$ where λ is observed wavelength

it is not hard to see that the amount of information contained in the Lyman-α forest is enormous. Most of the above observational facts were obtained using spectroscopic identification and counting of absorption lines together with profile-fitting techniques. This is not the only way, and perhaps not even necessarily the best way, to study the forest. An alternative is to take the forest in its entirety and study its aggregate properties, such as the mean transmission (transmission is the observed intensity divided by emitted intensity). For examples of such approaches applied to the observational data, see Jenkins and Ostriker (1991) and Press, Rybicki and Schneider (1993) and references therein. Moreover one can in principle study the variations in the transmission without restricting oneself to spectroscopically identified lines (Miralda-Escudé et al. 1995).

On the theoretical side, there exist several models of the Lyman-α forest. Traditionally, each absorption line is thought to arise from a cloud embedded in an otherwise smooth medium. Examples of such models include the Self-gravitating Clouds (Black 1981), the Pressure-confined Clouds (Ostriker & Ikeuchi 1983) and the Freely-expanding Clouds (Duncan et al. 1991). A discussion of them, in addition to other examples, can be found in the review article by Bajtlik (1993). Rees (1986) proposed that the gravity of dark matter mini-halos can confine intergalactic clouds. Bond, Szalay and Silk (1988) proposed a similar model based on the Cold Dark Matter (CDM) theory in which some of the clouds, in addition to being attracted by dark matter halos, were also allowed to expand when triggered by the reionization radiation. These were the first attempts to place the study of the Lyman-α forest within the larger framework of structure formation although they still subscribed to the traditional picture of distinct clouds embedded in an otherwise smooth medium. In subsequent work by McGill (1990) and Bi, Börner and Chu (1992), the traditional picture was given up in favor of one where the intergalactic medium had fluctuations on all scales due to initial conditions and gravitational instability, of which the Lyman-α forest was a manifestation. Recent work using numerical simulations gave a big boost to this way thinking by showing how its observational consequences could be computed in detail (Cen et al. 1994; Zhang et al. 1995; Hernquist et al. 1995; Petitjean

et al. 1995; Miralda-Escudé et al. 1995).

We adopt the same framework in this thesis. The question we would like to address is: what determines the shape and normalization of the column density distribution (the number of absorption lines per unit column density per unit redshift as a function of column density)? A set of tools is developed to calculate the column density distribution given a cosmological model and the thermal and ionization properties of the intergalactic medium. We make use of the Zel'dovich approximation to compute the density and velocity fields, thereby avoiding the use of full hydrodynamic simulations and allowing us to systematically study how the column density distribution depends on various cosmological parameters. In a sense, our work is a natural extension of the classic paper by McGill (1990) who simulated absorption spectra using one-dimensional Zel'dovich approximation.

1.4 Outline

The organization of the thesis is as follows. Chapters 2 and 3 cover the work on density fluctuations in the early universe. In Chapter 2 we lay down the groundwork and in Chapter 3, the main issue, the behavior of ζ close to the point of turn-around, is discussed. Chapters 4 and 5 cover the local approximations. In Chapter 4, we discuss the basic concepts of local approximations and introduce three examples. In Chapter 5, they are tested in the case of the collapse of a homogeneous ellipsoid. In Chapter 6, we apply the Zel'dovich approximation to compute the column density distribution of the Lyman-α forest. Finally, we briefly summarize the main results of the thesis in Chapter 7. Appendix A contains a summary of results from general relativistic perturbation theory that are used in Chapters 2 and 3. Appendix B and Appendix app-mij contains discussions of topics left over from Chapters 3 and 5. A list of symbols used in this thesis and their definitions can be found in Appendix D.

A word on the notation. Greek indices denote spatial-temporal components and Latin indices are reserved for spatial components. c, the speed of light, is always set to one, unless stated otherwise, in Chapter 6 for instance. Factors of G are kept

in Chapters 4 and 5 while $8\pi G$, as well as c, are set to one in Chapters 2 and 3.

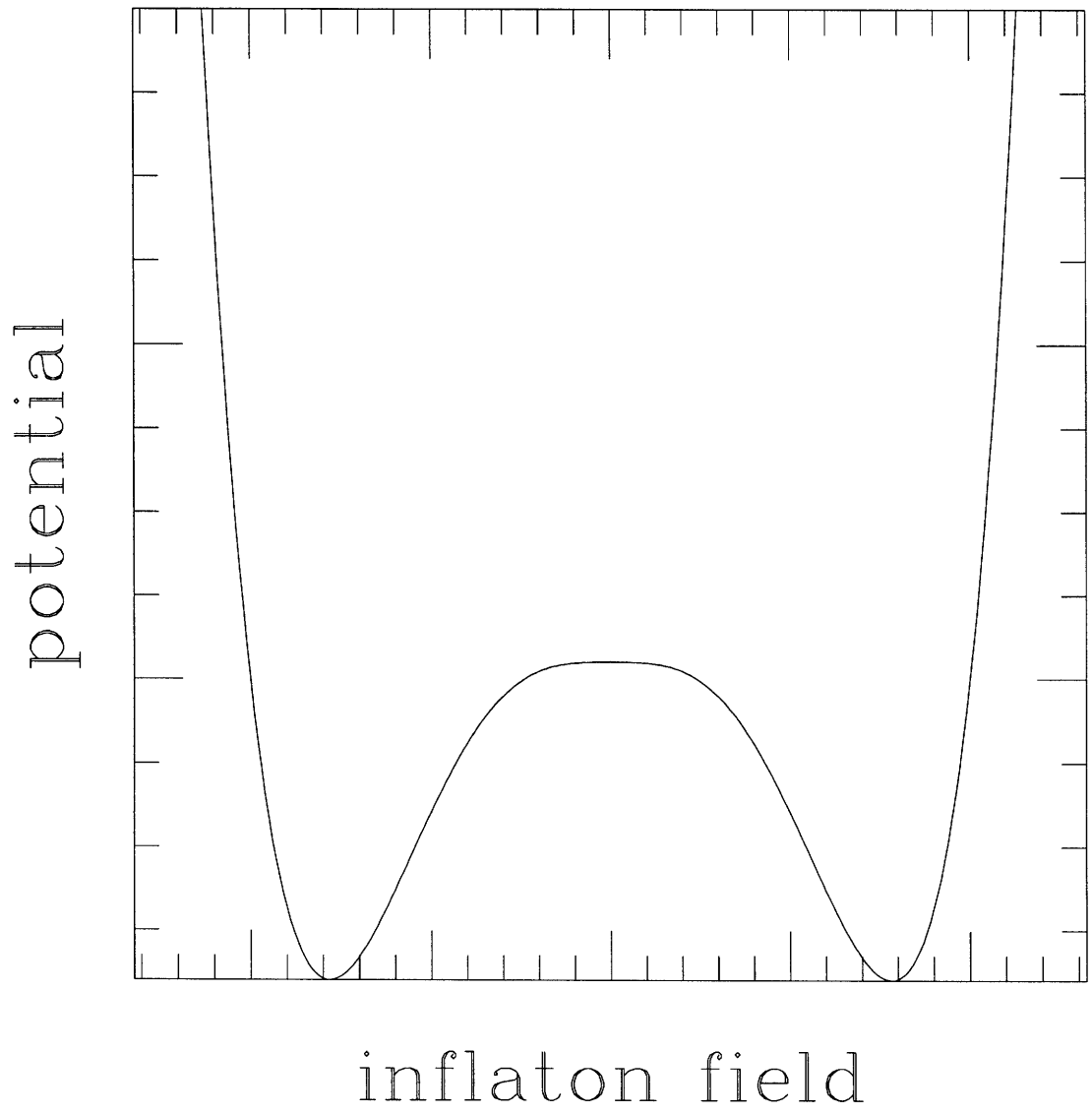

Fig. 1-1.— Potential of the Inflaton

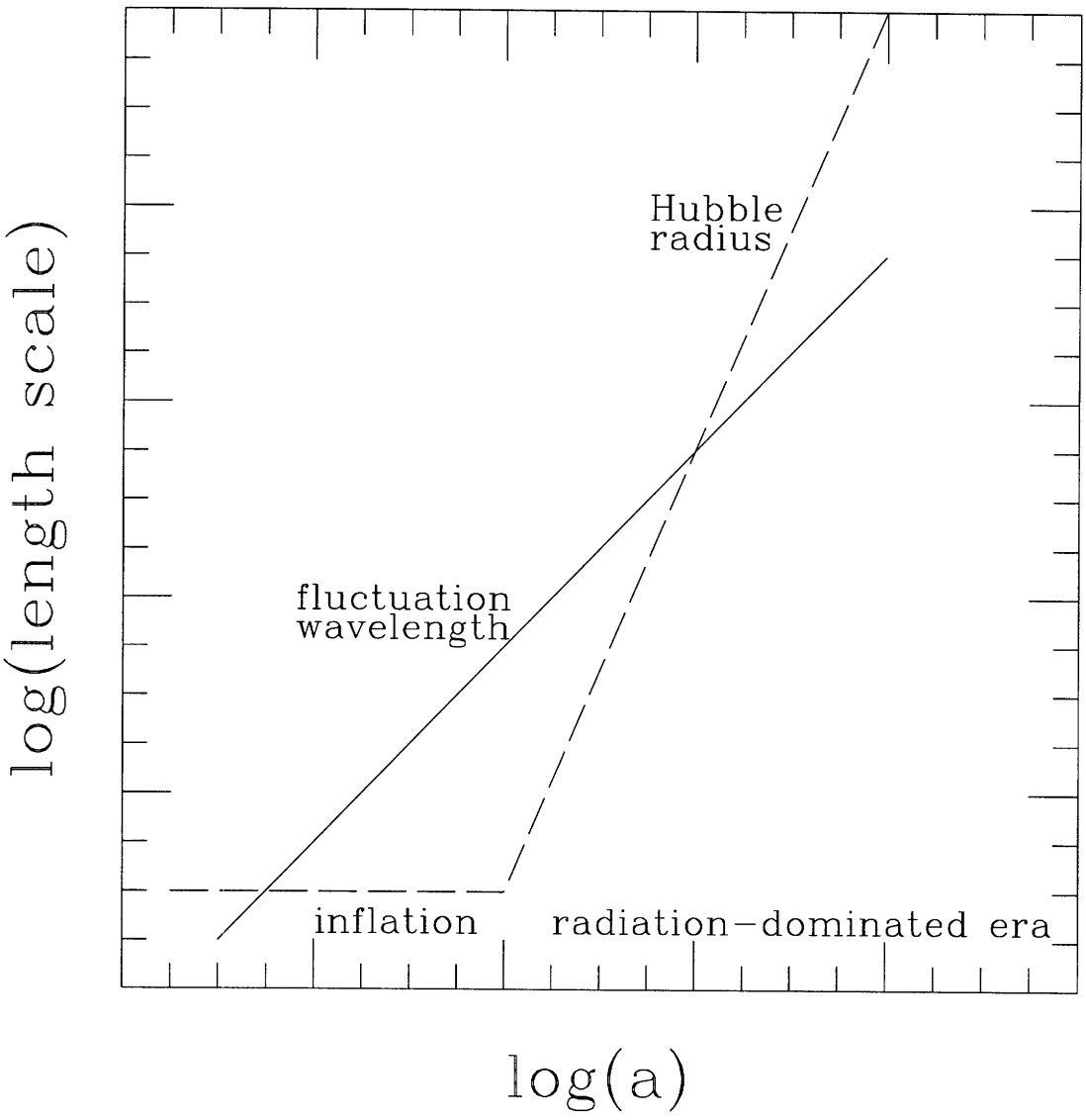

Fig. 1-2.— Growth of fluctuation length scale versus Hubble radius

Chapter 2

Density Fluctuations in the Early Universe: Basic Results

2.1 Introduction

The following expression has become part of the standard lore of the inflationary theory of cosmological fluctuations (Guth & Pi 1982; Bardeen et al. 1983; Starobinsky 1982; Hawking 1982):

$$\frac{\delta\rho(k)}{\rho_0} \propto \frac{H\delta\varphi(k)}{\dot{\varphi}_0}. \tag{2.1}$$

$\delta\rho(k)/\rho_0$ is the Fourier component (wave-number k) of a density fluctuation as it comes within the Hubble radius in the radiation-dominated or matter-dominated era. $\delta\varphi(k)$ is the fluctuation of the inflaton field. $\dot{\varphi}_0$ is the time derivative of the smooth background field and H is the Hubble parameter. The combination $H\delta\varphi(k)/\dot{\varphi}_0$ is evaluated at the time when the fluctuation exits the Hubble radius during inflation.

The density fluctuation is imprinted at the moment of Hubble-radius-exit. Its evolution outside the Hubble radius depends on the particular gauge one uses. But it is often argued that the amplitude of appropriately chosen gauge invariant combinations of metric and matter perturbations is frozen between the exit and re-entry of the Hubble radius. Typically their time derivatives are proportional to the spatial gradients of some other perturbed quantities. An example is the variable ζ,

which is related to the inflaton fluctuation by

$$\zeta = \frac{H\delta\varphi_{GI}}{\dot{\varphi}_0} \tag{2.2}$$

where $\delta\varphi_{GI}$ denotes a gauge invariant measure of $\delta\varphi$. ζ obeys the following simple equation

$$\frac{d\zeta}{dt} = -\frac{2\rho_0}{3(\rho_0 + p_0)}\frac{k^2}{Ha^2}\Phi_p \tag{2.3}$$

where ρ_0 and p_0 are the background density and pressure, a is the Hubble scale factor and Φ_p is a gauge invariant measure of pressure perturbation. These equations will be derived in the following sections. What is important is that in the long wavelength limit (small k), $\dot{\zeta}$ is negligible. This is often used as a justification that the details of reheating have no effects on the amplitude of density fluctuations at late times, since wavelengths of cosmological interests are much larger than the Hubble radius when reheating takes place. That is why in equation (2.1), the right hand side can be evaluated at the moment of Hubble-radius-exit: essentially the imprint of density fluctuation is unchanged by whatever happens between the exit and re-entry of the Hubble radius.

However, there is a factor of $\rho_0 + p_0$ in the denominator of the expression for $\dot{\zeta}$. The sum of density and pressure can exactly vanish. When the universe is inflaton-dominated, $\rho_0 + p_0$ is simply proportional to the kinetic energy of the scalar field. At the point of turn-around, when the inflaton has rolled past the minimum of the potential, decelerated and stops momentarily, the sum of energy and pressure vanishes and the argument for the conservation of ζ breaks down.

We examine the consequences in Chapters 2 and 3. An approximation of instantaneous transition is used. This is not meant to be realistic but it serves the purpose of demonstrating the possibilities. We calculate the evolution of density fluctuations for two cases: one where the inflationary stage ends during the slow-roll phase and one where it ends very close to the point of turn-around. The evolution of density fluctuations is followed only up to the radiation-dominated stage. It is straightforward to extend the analysis to the matter-dominated stage.

The organization of this chapter is as follows. The notation and description of the background and perturbed metric tensor and energy-momentum tensor is established in the next section. The evolution of gauge invariant metric variables in the inflationary and radiation-dominated era is discussed. In Section 2.3, We derive junction conditions for metric variables at the transition between the two stages. This is also discussed in Ratra (1991) and Deruelle (1995). Armed with the results in this chapter, we will be ready to tackle the problem of the evolution of ζ in the next. In order not to disrupt the flow of our arguments, we defer the proofs of a number of basic results in general relativistic perturbation theory to Appendix A.

A word on our terminology. When we use the word inflation, we mean the whole period during which the inflaton dominates the energy of the universe. This includes the slow-roll phase, as well as some of the initial reheating phase. In particular, in the framework of an instantaneous transition, by inflation we refer to the whole period before the transition to the radiation dominated stage. Immediately before the transition, the scalar field is not necessarily in the slow-roll phase. It depends on the model of the transition.

Both Chapters 2 and 3 are based on a preprint (Hui & Guth 1996). Greek indices denote spacetime components and Latin indices are reserved for space components. $c = 1$ and $8\pi G = 1$.

2.2 Formalism

The treatment follows closely that of Bardeen (1980) (see also Mukhanov, Feldman and Brandenberger 1992 and Grishchuk 1994). The key equations of this section are (2.17) and (2.19). Their counterparts are (2.28) and (2.29) in the inflationary stage and (2.33) and (2.34) in the radiation-dominated stage. The derivation of equation (2.17), as well as the proof of gauge invariance of the variables employed, are given in Appendix A.

The background spacetime is described by the flat Robertson-Walker metric:

$$ds^2 = a^2(\tau)(-d^2\tau + \delta_{ij}dx^i dx^j)\,. \tag{2.4}$$

The proper time t is related to the conformal time τ by $dt \equiv a d\tau$. We denote differentiation with respect to proper time by a dot and that with respect to the conformal time by a prime.

The unperturbed energy-momentum tensor is that of a perfect fluid at rest with respect to the above coordinates

$$T^0_0 = -\rho_0\,, \quad T^i_j = p_0\delta^i_j\,. \tag{2.5}$$

Mixed components vanish.

The evolution of the scale factor a is governed by

$$3H^2 = \rho_0 \tag{2.6a}$$

$$-\left(3H^2 + 2\frac{H'}{a}\right) = p_0\,, \tag{2.6b}$$

where $H \equiv \dot{a}/a = a'/a^2$ is the Hubble parameter.

The conservation of energy can be derived from the above equations.

$$\rho_0' + 3aH(\rho_0 + p_0) = 0 \tag{2.7}$$

We define two useful quantities γ and c_s^2.

$$\gamma \equiv -\dot{H}/H^2 = \frac{3(\rho_0 + p_0)}{2\rho_0}\,. \tag{2.8}$$

where the dot denotes differentiation with respect to proper time t. γ vanishes in de Sitter space and is expected to be small during most of the inflationary stage. It

equals 2 during the radiation-dominated era when $p_0 = \frac{1}{3}\rho_0$.

$$c_s^2 \equiv \frac{p_0'}{\rho_0'} = -1 + \frac{2}{3}\gamma - \frac{\gamma'}{3aH\gamma} \tag{2.9}$$

c_s^2 is not restricted to be positive.

Only scalar perturbations are considered in this thesis. They can be expanded in terms of spatial harmonics $Q_k(\mathbf{x})$ which satisfy

$$Q_{,i,i} + k^2 Q = 0 \tag{2.10}$$

where i is summed from 1 to 3. $,i$ denotes ordinary differentiation with respect to the i-th spatial coordinate. k is the wave-number. Each k mode evolves independently of the others in first order perturbation theory. From now on we assume the separation into individual harmonics has been made. The label k of perturbed quantities will be suppressed for the rest of the discussion.

The full metric is

$$g_{00} \equiv -a^2(\tau)[1 + 2A(\tau)Q] \tag{2.11a}$$

$$g_{0i} \equiv a^2(\tau)B(\tau)\frac{Q,i}{k} \tag{2.11b}$$

$$g_{ij} \equiv a^2(\tau)[(1 + 2H_L(\tau)Q)\delta_{ij} + 2H_T(\tau)(k^{-2}Q_{,i,j} + \frac{1}{3}\delta_{ij}Q)]\,. \tag{2.11c}$$

The full energy-momentum tensor is

$$T_0^0 \equiv -(\rho_0 + \delta\rho(\tau)Q) \tag{2.12a}$$

$$T_0^i \equiv (\rho_0 + p_0)v(\tau)\frac{Q^{,i}}{k} \tag{2.12b}$$

$$T_i^0 \equiv -(\rho_0 + p_0)(v(\tau) - B(\tau))\frac{Q_{,i}}{k} \tag{2.12c}$$

$$T_j^i \equiv (p_0 + \delta p(\tau)Q)\delta_j^i\,. \tag{2.12d}$$

It is implicitly assumed that the anisotropic stress (traceless part of T_j^i) is negligible.

$Q^{,i}$ and $Q_{,i}$ denote the same ordinary derivative of Q with respect to the spatial coordinate. Note that the perturbation amplitudes A, B, H_L, H_T, v, $\delta\rho$ and δp are all functions of k as well as τ. We define the following gauge invariant measures of the density and pressure perturbations:

$$\delta\rho_{GI} \equiv \delta\rho + 3\frac{aH}{k}(\rho_0 + p_0)(v - B) \tag{2.13a}$$

$$\delta p_{GI} \equiv \delta p + 3\frac{aH}{k}(\rho_0 + p_0){c_s}^2(v - B)\,. \tag{2.13b}$$

They exactly equal $\delta\rho$ and δp respectively in the comoving gauge where $T_i^0 = 0$. Their gauge invariance is proved in Appendix A.

Two gauge invariant combinations of the metric perturbations will be used throughout this and the next chapter:

$$\Phi \equiv \frac{aH}{k^2}(H_T' - kB) - (H_L + \frac{1}{3}H_T) \tag{2.14a}$$

$$\zeta \equiv \frac{\Phi'(aH)^{-1} + \Phi}{\gamma} + \Phi\,, \tag{2.14b}$$

where γ is defined in equation (2.8). The definition of ζ follows that of Grishchuk (1992). There is an alternative, physically more transparent definition:

$$\zeta \equiv \frac{3a^2}{2k^2}\left(H_{\rm local}^2 - \frac{\rho}{3}\right) \tag{2.15}$$

where ρ is the full density and $H_{\rm local} \equiv \nabla_\alpha U^\alpha/3$, where ∇_α denotes covariant differentiation. U^α is the fluid velocity. For perfect fluids, which we consider in this thesis, $T^{\alpha\beta} = (\rho+p)U^\alpha U^\beta + pg^{\alpha\beta}$. $H_{\rm local}$ is a measure of the local expansion rate with respect to the fluid observers. The uniform part of $H_{\rm local}$ is just the Hubble constant. So for a flat universe, the combination $H_{\rm local}^2 - \frac{\rho}{3}$ has no homogeneous part. The quantity ζ is then a measure of how much the universe departs from flatness locally. It will be shown in Appendix A that the two definitions, equations (2.14b) and (2.15), are equivalent. Note also that equation (2.15) is expressed in Fourier space and we have dropped the spatial harmonics Q.

It is shown in Appendix A that Φ and $\delta\rho_{GI}$ are related by the Poisson equation:

$$\Phi = -\frac{a^2}{2k^2}\delta\rho_{GI}\,. \tag{2.16}$$

We also prove in Appendix A that Φ satisfies the following equation (equation [4.9] of Bardeen (1980)):

$$\frac{d^2\Phi}{dt^2} + \left[H - \frac{(\rho_0+p_0)^\bullet}{\rho_0+p_0}\right]\frac{d\Phi}{dt} - H\frac{(\rho_0+p_0)^\bullet}{\rho_0+p_0}\Phi = (\rho_0+p_0)\Phi - \frac{k^2}{a^2}\Phi_p\,, \tag{2.17}$$

where the dot denotes differentiation with respect to proper time. Φ_p is defined by analogy with equation (2.16).

$$\Phi_p \equiv -\frac{a^2}{2k^2}\delta p_{GI} \tag{2.18}$$

It follows from the definition of ζ (equation [2.14b]) that

$$\frac{d\zeta}{dt} = -\frac{2\rho_0}{3a^2H(\rho_0+p_0)}k^2\Phi_p \tag{2.19}$$

Equations (2.17) and (2.19) are completely general. The only assumption made is that the anisotropic stress and the background spatial curvature is negligible. A similar equation was derived by Mukhanov, Feldman and Brandenberger (1992) but the right hand side was put to zero at some point in their derivation assuming the small k limit. It is commonly argued based on equation (2.19) that ζ is roughly constant when k^2 or k^2/a^2H^2 is small enough (the long wavelength limit). This is often taken to imply that the amplitude of density fluctuations (which is related to the amplitude of ζ) is unaffected by reheating because wavelengths of cosmological interests are much longer than the Hubble radius at the end of inflation. We will discuss where this argument breaks down.

2.2.1 The Inflationary Era

The energy density of the universe is dominated by that associated with a scalar field $\varphi(\tau, \mathbf{x})$ which can be split into two parts, a smooth background and a fluctuation:

$$\varphi(\tau, \mathbf{x}) = \varphi_0(\tau) + \delta\varphi(\tau)Q(\mathbf{x})\,. \tag{2.20}$$

The energy-momentum tensor is

$$T_{\mu\nu} = \varphi_{,\mu}\varphi_{,\nu} - g_{\mu\nu}\left[\frac{1}{2}g^{\alpha\beta}\varphi_{,\alpha}\varphi_{,\beta} + V(\varphi)\right] \tag{2.21}$$

where $V(\varphi)$ is the potential of the inflaton.

Hence the background density and pressure, ignoring contributions from matter or radiation, are

$$\rho_0 = \frac{1}{2a^2}\varphi_0'^{\,2} + V(\varphi) \tag{2.22a}$$

$$p_0 = \frac{1}{2a^2}\varphi_0'^{\,2} - V(\varphi)\,. \tag{2.22b}$$

The equation of motion for φ_0 follows from the conservation of energy (equation [2.7]).

$$\varphi_0'' + 2aH\varphi_0' + a^2\left.\frac{\partial V}{\partial\varphi}\right|_{\varphi=\varphi_0} = 0\,. \tag{2.23}$$

The perturbations in density and pressure are

$$\delta\rho = -\frac{A}{a^2}\varphi_0'^{\,2} + \frac{1}{a^2}\varphi_0'\delta\varphi' + \left.\frac{\partial V}{\partial\varphi}\right|_{\varphi=\varphi_0}\delta\varphi \tag{2.24a}$$

$$\delta p = -\frac{A}{a^2}\varphi_0'^{\,2} + \frac{1}{a^2}\varphi_0'\delta\varphi' - \left.\frac{\partial V}{\partial\varphi}\right|_{\varphi=\varphi_0}\delta\varphi \tag{2.24b}$$

where A is the metric perturbation defined in equation (2.11a).

The scalar field fluctuation is also related to the energy flux (equation (2.12c).

$$T_i^0 = -\frac{\varphi_0'}{a^2}\delta\varphi Q_{,i} = -\frac{1}{k}(\rho_0 + p_0)(v(\tau) - B(\tau))Q_{,i} \tag{2.25}$$

Let me define a gauge invariant measure of the inflaton field fluctuation:

$$\delta\varphi_{GI} \equiv \delta\varphi - \frac{\varphi_0'}{aH}(H_L + \frac{1}{3}H_T) \tag{2.26}$$

where H_L and H_T are the metric perturbations defined in equation (2.11). The proof of its gauge invariance is left to Appendix A. It will also be shown that that

$$\delta\varphi_{GI} = \frac{\varphi_0'}{aH}\zeta \tag{2.27}$$

where ζ is defined in equation (2.14b).

Equation (2.17) becomes a very useful equation in the inflationary era because $\Phi_p = \Phi$. To see this is the case, first note that both Φ_p and Φ are gauge invariant. It is sufficient to demonstrate their equivalence by proving so in one particular gauge. In the comoving gauge where $T_i^0 = 0$ and hence $\delta\varphi = 0$ (equation [2.25]), $\delta\rho$ equals δp by equation (2.24). It then follows from equation (2.13) that $\delta\rho_{GI} = \delta p_{GI}$. Therefore $\Phi_p = \Phi$ (equations [2.16] and [2.18]).

Substitute this relation into equations (2.17) and (2.19), the following is obtained:

$$\frac{d^2\Phi}{dt^2} + \left(H - \frac{(\rho_0 + p_0)^\bullet}{\rho_0 + p_0}\right)\frac{d\Phi}{dt} - H\frac{(\rho_0 + p_0)^\bullet}{\rho_0 + p_0}\Phi = (\rho_0 + p_0)\Phi - \frac{k^2}{a^2}\Phi \tag{2.28}$$

and

$$\frac{d\zeta}{dt} = -\frac{2\rho_0}{3a^2H(\rho_0 + p_0)}k^2\Phi \tag{2.29}$$

Making use of the above equations and the definition of ζ (equation [2.14b]), a single equation for ζ can be written down:

$$\frac{d^2\zeta}{dt^2} + \left(3H + \frac{\dot{\gamma}}{\gamma}\right)\frac{d\zeta}{dt} + \frac{k^2}{a^2}\zeta = 0 \tag{2.30}$$

where γ is defined in equation (2.8). The fact that $\rho_0 + p_0 = \dot{\varphi}_0^2$, together with equation (2.6a), implies that

$$\gamma = \frac{{\varphi_0'}^2}{2a^2H^2} \tag{2.31}$$

γ is a measure of the contribution of kinetic energy to the total energy content of the universe.

Equation (2.30) can be rewritten as:

$$(a\sqrt{\gamma}\zeta)'' + \left(k^2 - \frac{(a\sqrt{\gamma})''}{a\sqrt{\gamma}}\right)(a\sqrt{\gamma}\zeta) = 0 \tag{2.32}$$

where the prime stands for differentiation with respect to the conformal time τ. $\sqrt{\gamma} = \varphi_0'/\sqrt{2}aH$ and can be positive or negative depending on the sign of φ_0'.

In this thesis we adopt the approximation that the inflationary era ends abruptly, giving rise to a radiation-dominated universe. It is worth emphasizing that by the inflationary stage we include not just the slow-roll phase, but, depending on when we choose inflation to end, possibly also the phase when the scalar field rolls quickly and oscillates around the minimum of the potential. The potential and kinetic energy associated with the scalar field dominates the energy content of the universe until the next stage, the radiation-dominated era.

2.2.2 The Radiation-dominated Era

In the radiation-dominated universe, $p_0 = \frac{1}{3}\rho_0$ and $\delta p = \frac{1}{3}\rho$. There is no entropy perturbation. Equation (2.7) implies that ρ evolves like a^{-4} from which it follows that $a \propto t^{1/2} \propto \tau$ by equation (2.6a).

In this phase $\Phi_p = \frac{1}{3}\Phi$. We can show it in the comoving gauge, where $\delta\rho_{GI} = \delta\rho$ and $\delta p_{GI} = \delta p$ (equation [2.13]). Since $\delta p = \frac{1}{3}\rho$, equations (2.16) and (2.18) imply that $\Phi_p = \frac{1}{3}\Phi$. It then follows from equations (2.17) and (2.19) that

$$\frac{d^2\Phi}{dt^2} + \left(H - \frac{(\rho_0 + p_0)^\bullet}{\rho_0 + p_0}\right)\frac{d\Phi}{dt} - H\frac{(\rho_0 + p_0)^\bullet}{\rho_0 + p_0}\Phi = (\rho_0 + p_0)\Phi - \frac{k^2}{a^2}\frac{\Phi}{3} \tag{2.33}$$

and

$$\frac{d\zeta}{dt} = -\frac{2\rho_0}{9a^2H(\rho_0+p_0)}k^2\Phi\,. \tag{2.34}$$

Note that in the radiation-dominated universe the above equations simplify enormously because $\rho_0 + p_0 = 4\rho_0/3$ and $(\rho_0+p_0)^\bullet/(\rho_0+p_0) = -4H$. A particularly simple equation for $a\zeta$ can be written down using equations (2.33), (2.34), (2.14b) and the fact that $\gamma \equiv -\dot{H}/H = 2$ (equation [2.8]) in the radiation-dominated universe.

$$(a\zeta)'' + \frac{k^2}{3}(a\zeta) = 0 \tag{2.35}$$

where the prime denotes differentiation with respect to conformal time τ.

An explicit solution for ζ is

$$\zeta = \frac{K_{r1}}{a}\sin(\frac{k\tau}{\sqrt{3}}) + \frac{K_{r2}}{a}\cos(\frac{k\tau}{\sqrt{3}}) \tag{2.36}$$

where K_{r1} and K_{r2} are constants of integration. Equation (2.34) then implies

$$\Phi = 2a^{-1}(\frac{k\tau}{\sqrt{3}})^{-2}K_{r1}\left[\sin(\frac{k\tau}{\sqrt{3}}) - \frac{k\tau}{\sqrt{3}}\cos(\frac{k\tau}{\sqrt{3}})\right] + 2a^{-1}(\frac{k\tau}{\sqrt{3}})^{-2}K_{r2}\left[\cos(\frac{k\tau}{\sqrt{3}}) + \frac{k\tau}{\sqrt{3}}\sin(\frac{k\tau}{\sqrt{3}})\right]. \tag{2.37}$$

We have used $aH = 1/\tau$.

Relating Φ to $\delta\rho_{GI}$ through equation (2.16), it can be shown that $\delta\rho_{GI}/\rho_0$ oscillates with a nearly constant amplitude after the fluctuation comes within the Hubble radius ($k/aH \gg 1$). In this limit, measurements of the density fluctuation in different coordinate systems agree with each other. $\delta\rho/\rho_0$ in this regime is given by:

$$\frac{\delta\rho}{\rho_0} = \frac{4k}{\sqrt{3}a^2H}\left[K_{r1}\cos(\frac{k\tau}{\sqrt{3}}) - K_{r2}\sin(\frac{k\tau}{\sqrt{3}})\right], \tag{2.38}$$

where $4k/(\sqrt{3}a^2H)$ is a constant in time because $a \propto \tau$. The amplitude of oscillation is determined by the constants K_{r1} and K_{r2} which are evaluated by matching ζ and Φ across the transition between the inflationary and radiation-dominated stages. This is the subject of the next section.

2.3 JUNCTION CONDITIONS

The inflationary stage is assumed to end abruptly at an equal energy density surface (Ratra 1991; Deruelle & Mukhanov 1995). Inflation allows small scale quantum fluctuations in φ and hence in ρ to expand to large scales. ρ then acts as a timer: when it reaches a particular value, inflation ends, reheating takes place instantaneously and gives rise to a radiation-dominated universe. To follow the evolution of fluctuations, we need junction conditions for ζ and Φ, which satisfy equations (2.17) and (2.19).

First we will show that Φ is continuous across the transition. This will be demonstrated in the synchronous gauge. The gauge invariance of Φ guarantees that it is continuous in any other coordinate system.

In the synchronous gauge, $A = B = 0$ in equation (2.11). We define the following symbols for simplicity of notation:

$$h \equiv 2(H_L + \frac{1}{3}H_T) \quad , \quad h_\ell \equiv 2H_T \tag{2.39}$$

In these variables, the first order Einstein Equations are

$$\frac{3H}{a}h' + \frac{k^2}{a^2}h - \frac{H}{a}h'_\ell = \delta\rho \tag{2.40a}$$

$$h' = -\frac{a^2(\rho_0 + p_0)}{k}v \tag{2.40b}$$

$$-\frac{1}{a^2}(h'' + 2aHh') = \delta p \tag{2.40c}$$

$$h''_\ell + 2aHh'_\ell - k^2 h = 0\,. \tag{2.40d}$$

The symbols on the right hand side are defined in equation (2.12). The first two follow from the G^0_0 and G^0_i equations. The last two are derived from the G^i_j equations. The last equation is a result of the no-anisotropic-stress assumption.

First note that a and H are continuous by virtue of equation (2.6) and the fact that neither ρ_0 nor p_0 has a delta function at the transition. In fact, ρ_0 is continuous while p_0 has a discontinuity at the transition. One has the freedom in the synchronous gauge to choose the transition surface (equal-energy-density surface) to be an equal-

time surface. In this coordinate system, $\delta\rho = 0$ and there are no delta functions on the right hand side of the above four equations. It automatically follows that h, h_ℓ, h' and h'_ℓ are continuous which implies $\delta\rho$ itself is continuous. δp, on the other hand, is allowed to have a jump at the transition.

It is worth pointing out that the above conclusions do not hold if one uses a coordinate system in which the transition (equal-energy-density) surface is not an equal-time surface. Suppose a new synchronous coordinate system is related to the one used in the last paragraph by $\tau_{new} = \tau_{old} + \Delta\tau(\mathbf{k})\mathbf{Q}(\mathbf{x})$ (together with a corresponding change in $\mathbf{x}$). Then $\delta p_{new} = \delta p_{old} - p'_0\Delta\tau$. Since the background pressure p_0 is discontinuous at the transition, δp_{new} contains a delta function and equation (2.40c) implies that h' is not continuous.

This does not mean that the calculation cannot be carried out in the new coordinate system. As long as the appropriate discontinuity in the derivatives of the metric are taken into account, any coordinate system can be used. One can however make a gauge independent statement regarding the junction condition. Φ, which equals $(aHh'_\ell - k^2h)/(2k^2)$ in the synchronous gauge by equations (2.14a) and (2.39), is continuous across the transition. This is because h and h'_ℓ are continuous in the particular synchronous coordinate system in which the transition surface is an equal-time surface. The gauge invariance of Φ then guarantees its continuity in any coordinate system.

One more junction condition is needed because we have to match solutions of a second order differential equation (2.17) at the transition. We will derive the discontinuity in ζ (equation [2.19]) as the second junction condition.

Φ_p, defined in equation (2.18), is proportional to δp_{GI}. According to equation (2.13b), δp_{GI} contains a delta function due to the presence of c_s^2. This is because p_0 or γ jumps at the transition (equation [2.9]).

To derive the jump in ζ, let me again employ the synchronous coordinate system in which the transition (equal-energy-density) surface is an equal-time surface. At the transition, $\delta\rho = 0$ implies that $h' = -(k^2h - aHh'_\ell)/3aH$ (equation [2.40a]). From equation (2.40b), it follows that $-a^2(\rho_0 + p_0)v/k = -(k^2h - aHh'_\ell)/3aH$. The

part of δp_{GI} that contains a delta function is therefore $c_s^2(k^2 h - aHh'_\ell)/a^2$. Note that δp does not have a delta function because we are working in coordinates in which the transition surface is an equal-time surface. Using equations (2.9), (2.18) and (2.19), it can be verified that ζ has a jump at the transition:

$$\Delta\zeta = -\frac{k^2\Phi}{3a^2H^2}\Delta\left(\frac{1}{\gamma}\right) \tag{2.41}$$

where Δ denotes the change across the transition. We have used the fact that $\Phi = (aHh'_\ell - k^2 h)/(2k^2)$ in the synchronous coordinate system.

A new gauge invariant quantity can be defined,

$$\zeta_\rho \equiv \zeta + \frac{k^2\Phi}{3\gamma a^2 H^2} \tag{2.42}$$

which is obviously continuous in the particular synchronous gauge used and so is continuous in any coordinate system.

The continuity of ζ_ρ or equation (2.41) can be used as the second junction condition. Note that because of the jump, ζ does not remain unchanged through the transition. However, the jump is very small if the inflationary stage ends during the slow-roll phase. We will discuss this point in the next chapter.

Let me denote the values of Φ and ζ right before and after the transition by the subscripts infl and rad. Let me also define the following symbols:

$$x_{\rm tr} \equiv \frac{k\tau_{\rm tr}}{\sqrt{3}} = \frac{k}{\sqrt{3}a_{\rm tr}H_{\rm tr}} \tag{2.43a}$$

$$F \equiv \zeta_{\rm infl} - x_{\rm tr}^2\Phi_{\rm infl}\Delta\left(\frac{1}{\gamma}\right) \tag{2.43b}$$

$$G \equiv \frac{x_{\rm tr}^2}{2}\Phi_{\rm infl} \tag{2.43c}$$

where the subscript tr denotes evaluation at the transition. $\Delta(1/\gamma) = 1/\gamma_{\rm rad} - 1/\gamma_{\rm infl}$.

Then the continuity of Φ and the discontinuity of ζ (equation [2.41]), together

with equations (2.36) and (2.37) implies that

$$K_{r1} = (F - G)a_{\rm tr}\frac{\cos x_{\rm tr}}{x_{\rm tr}} + Fa_{\rm tr}\sin x_{\rm tr} \tag{2.44a}$$

$$K_{r2} = -(F - G)a_{\rm tr}\frac{\sin x_{\rm tr}}{x_{\rm tr}} + Fa_{\rm tr}\cos x_{\rm tr} \tag{2.44b}$$

Fluctuations of cosmological interest have wavelengths much longer than the Hubble radius at the time of transition, hence $x_{\rm tr} \ll 1$. In this limit, K_{r1} is approximately equal to $(F-G)a_{\rm tr}/x_{\rm tr}$ and K_{r2} is insignificant compared to K_{r1}. Substituting this into equation (2.38), we obtain the asymptotic amplitude of acoustic oscillation after the fluctuation enters the Hubble radius in the radiation-dominated phase:

$$\left.\frac{\delta\rho}{\rho_0}\right|_{\rm asymp} = 4(F - G) \tag{2.45}$$

where F and G are defined in equations (2.43b) and (2.43c). We have used the definition of $x_{\rm tr}$ (equation [2.43a]) and the fact that $a^2H = a_{\rm tr}^2 H_{\rm tr}$ in the radiation-dominated stage.

The amplitude of the acoustic oscillation is completely determined by the values of ζ and Φ immediately before the transition. It is worth emphasizing that Φ is related to the derivative of ζ through equation (2.29). Thus it is sufficient to follow the evolution of ζ in order to evaluate F and G in equation (2.45).

Chapter 3

Density Fluctuations in the Early Universe: Dependence on Reheating

3.1 Introduction

In the last chapter, we have reviewed the basic results of general relativistic linear perturbation theory, for the inflationary and radiation dominated stages. It has been shown that the amplitude of density fluctuations at second Hubble-radius-crossing during the radiation-dominated era is determined by the values of ζ and Φ at the end of inflation (equation [2.45]). Our aim in the chapter is to find out how these values depend on the manner in which inflation ends.

The next section is devoted to a study of the evolution of the important quantity ζ in the inflationary era. In Section 3.3, the singularity of ζ at the point of turn-around is analyzed. In Section 3.4, we put everything together and derive the amplitude of density fluctuations in the radiation-dominated universe for two different models of the transition, one where instantaneous reheating occurs in the slow-roll phase and the other where it takes place very close to the point of turn-around. We conclude in Section 3.5. In Appendix B to this chapter, we revisit the behavior of ζ close to the point of turn-around and reconcile it with the interpretation of ζ as a curvature fluctuation.

3.2 The Evolution of ζ in the Inflationary Era

The evolution of ζ is determined by the simple equation (2.32). There are clearly two separate regimes. When $k^2 \gg (a\sqrt{\gamma})''/a\sqrt{\gamma}$, $a\sqrt{\gamma}\zeta$ simply oscillates. Hence,

$$\zeta = \frac{\sqrt{2}H}{\varphi_0'}[K_{q1}\sin(k\tau) + K_{q2}\cos(k\tau)] \tag{3.1}$$

where K_{q1} and K_{q2} are integration constants. They are determined by the quantization procedure, which will not be discussed in the present work (see Mukhanov, Feldman and Brandenberger 1992). Note that we have used $\sqrt{\gamma} = \varphi_0'/\sqrt{2}aH$ (equation [2.31]), which can be positive or negative.

The other regime is when $k^2 \ll (a\sqrt{\gamma})''/a\sqrt{\gamma}$. This is further divided into two periods, a plateau/slow-roll phase and a valley phase. By the plateau we refer to the stage when the scalar field slowly rolls at the top of the potential. The energy content of the universe is dominated by the potential energy associated with the scalar field. The Hubble parameter H is nearly constant and the universe expands almost exponentially. The subsequent phase, when the scalar field departs significantly from the plateau, is called the valley phase. The scalar field is not in the flat part of the potential in this phase.

In the plateau or slow-roll phase, the kinetic energy of the inflaton is an insignificant portion of the total energy, hence $\dot{\varphi}_0^2 \ll H^2$. The acceleration is also very small, which means $\ddot{\varphi} \ll H\dot{\varphi}_0$. If we further assume $d^3\varphi_0/dt^3 \ll H^2\dot{\varphi}_0$, it can be shown that $(a\sqrt{\gamma})''/a\sqrt{\gamma} \sim 2a^2H^2$ by using the expression for γ in equation (2.31). Hence the transition between the early oscillatory phase (equation [3.1]) and the plateau/slow-roll phase ($k^2 \ll (a\sqrt{\gamma})''/a\sqrt{\gamma} \sim 2a^2H^2$) corresponds to Hubble-radius-crossing.

An approximate solution for equation (2.32) in both the plateau/slow-roll phase and the valley phase is

$$\zeta(t) = K_1 + K_2 \int_{t_{cr}}^{t} \frac{d\bar{t}}{a^3\gamma} \tag{3.2}$$

where we use proper time t instead of conformal time τ and $t_{\rm cr}$ is the time of Hubble-radius-crossing. K_1 and K_2 are integration constants which can be approximately related to K_{q1} and K_{q2} (equation [3.1]) by matching ζ and its derivative at $t_{\rm cr}$.

Note that $\dot{\zeta} = K_2/\gamma a^3$. γ, which is equal to $\dot{\varphi}_0^2/2H^2$ (equation [2.31]), increases with time in the plateau/slow-roll phase. That H decreases with time follows from $\dot{H} = -\dot{\varphi}_0^2/2$ (equation [2.8]). The time dependence of $\dot{\varphi}_0^2$ follows from the slow-roll approximation that $\dot{\varphi}_0 \sim -(3H)^{-1}\partial V/\partial\varphi$ (by rewriting equation [2.23] using proper time and ignoring the $\ddot{\varphi}_0$ term). The factor of H^{-1} increases with time as argued before. If one assumes the potential is convex (i.e. $\partial^2 V/\partial\varphi^2 < 0$), $(\partial V(\varphi_0)/\partial\varphi)^2$ also increases with time as φ_0 slowly moves away from zero or the top of the potential.

Hence both γ and a in the denominator of the expression for $\dot{\zeta}$ increase monotonically in the slow roll phase. a grows almost exponentially in this phase because the energy content of the universe is dominated by the potential energy of the inflaton, or a cosmological constant. The enormous expansion of a causes $\dot{\zeta}$ to decay quickly. ζ approaches a constant soon after $t_{\rm cr}$, the time of Hubble-radius-crossing.

This behavior is the origin of the statement that ζ remains constant for a fluctuation of long enough wavelength. This is commonly demonstrated through equation (2.29) or its cousins. Note that $\rho_0 + p_0$ or γ (equation [2.8]) occurs in the denominator of the expression for $\dot{\zeta}$. As a measure of the contribution of kinetic energy to the total energy content of the universe (equation [2.31]), γ is small but non-zero during the plateau/slow-roll phase. As long as k^2 is sufficiently small in the slow-roll phase, $\dot{\zeta}$ is negligible according to equation (2.29). However, γ or $\rho_0 + p_0$ can vanish in the valley phase, in which case the argument for the conservation of ζ breaks down. Let us now turn to the valley phase.

We denote by t_v the time when the plateau/slow-roll phase ends and the valley phase begins. Equation (3.2) is rewritten in the form:

$$\zeta(t) = \left(K_1 + K_2 \int_{t_{\rm cr}}^{t_v} \frac{d\bar{t}}{a^3\gamma}\right) + K_2 \int_{t_v}^{t} \frac{d\bar{t}}{a^3\gamma} \tag{3.3}$$

The terms in the bracket on the right hand side give the asymptotic value of ζ in the

plateau/slow-roll phase. ζ rapidly approaches this limit soon after Hubble-radius-crossing, as argued before. Let us define

$$\zeta_{\rm sl} \equiv K_1 + K_2 \int_{t_{\rm cr}}^{t_v} \frac{d\bar{t}}{a^3\gamma} \tag{3.4}$$

to represent this asymptotic value.

At the start of the valley phase, the inflaton begins picking up speed because it is now in the steeper part of the potential. Then it rolls past the true minimum of the potential and starts decelerating. Eventually, the inflaton comes to a halt, $\dot{\varphi}_0 = 0$. γ (equation [2.31]) vanishes. According to equation (3.3), $\zeta(t)$ diverges. I define $t = 0$ to be the instant when this occurs, the time of turn-around. It raises a few interesting questions. Does Φ, which is a measure of the density fluctuation (equation [2.16]), also diverge? Does ζ remain infinite after the inflaton starts moving again? These questions are addressed in the next section.

3.3 Singularity at the Turn-around

To understand the behavior of ζ and Φ close to the turn-around, we solve equations (2.28) and (2.30) by a series expansion. Let $t = 0$ be the instant of turn-around. We expand the coefficients of the differential equations in powers of t. At $t = 0$, $\dot{H} = -\dot{\varphi}_0^2/2 = 0$ (equations [2.8] and [2.31]). Hence $\ddot{H}$ vanishes at $t = 0$ also but the third derivative is non-zero if $\ddot{\varphi}_0 \neq 0$. Equation (2.30) can be rewritten in the form

$$\ddot{\zeta} + C(t)\dot{\zeta} + D(t)\zeta = 0 \tag{3.5}$$

where $C(t) = \sum_{n=-1}^{\infty} C^{(n)} t^n$ and $D(t) = \sum_{n=0}^{\infty} D^{(n)} t^n$. The first few coefficients are

$$C^{(-1)} = 2 \quad , \quad D^{(0)} = \left.\frac{k^2}{a^2}\right|_{t=0} \tag{3.6a}$$

Similarly, equation (2.28) can be rewritten as

$$\ddot{\Phi} + X(t)\dot{\Phi} + Y(t)\Phi = 0 \tag{3.7}$$

where $X(t) = \sum_{n=-1}^{\infty} X^{(n)} t^n$ and $Y(t) = \sum_{n=-1}^{\infty} Y^{(n)} t^n$. The lowest order coefficients are

$$X^{(-1)} = -2 \quad , \quad Y^{(-1)} = -2H|_{t=0} \tag{3.8a}$$

The series solutions for ζ and Φ are

$$\zeta = \quad (\zeta^I + C^{(0)}\zeta^{II} \ln|t|)\left[1 - \frac{D^{(0)}}{6}t^2 + ...\right] + \zeta^{II}\left[\frac{1}{t} + ...\right] \tag{3.9a}$$

$$\begin{aligned}\Phi = \quad & (\Phi^I + Z\Phi^{II} \ln|t|)\left[t^3 + ...\right] \\ & + \Phi^{II}\left[1 - H|_{t=0}\, t + \left(\frac{-HX^{(0)}}{2} + H^2 + \frac{Y^{(0)}}{2}\right)\Big|_{t=0} t^2 + ...\right]\end{aligned} \tag{3.9b}$$

where ζ^I, ζ^{II}, Φ^I and Φ^{II} are integration constants. Z is a finite constant that depends on the values of a and its derivatives at $t = 0$. Its exact form is unimportant for our purpose. Note that the constants ζ^I and ζ^{II} are not independent of Φ^I and Φ^{II} because ζ and Φ are related by equation (2.14b).

The above expressions can be obtained by assuming ζ (or Φ) has a solution of the form $t^m \sum_{n=0}^{\infty} b^{(n)} t^n$ where $b^{(n)}$'s are constants. This is substituted into the respective differential equations. m is determined by requiring that $b^{(0)}$ does not vanish. It turns out this method gives only one of the two independent solutions, the one associated with the constants ζ^I (or Φ^I). Let us denote this solution by $s^I(t)$. The other independent solution $s^{II}(t)$ is obtained by solving an equation for the Wronskian which is defined as

$$W \equiv s^I \dot{s}^{II} - \dot{s}^I s^{II} = (s^I)^2 \frac{d}{dt}\left(\frac{s^{II}}{s^I}\right) . \tag{3.10}$$

The equation for the W can be derived from the differential equation for ζ or Φ. $-\dot{W}/W$ is equal to $C(t)$ or $X(t)$ depending on whether one is solving equation (3.5) or (3.7).

The ζ expansion contains divergent terms proportional to $\ln|t|$ and t^{-1}. In the small t limit, t^{-1} dominates. Note that ζ does not remain infinite for $t > 0$ (after the point of turn-around). Therefore ζ is well-behaved after the inflaton starts moving again, even though it diverges at the instant of turn-around. Φ, on the other hand, is regular at $t = 0$. The divergence of $\ln|t|$ is killed by powers of t multiplying it. The gauge invariant density perturbation $\delta\rho_{GI}$, which is simply proportional to Φ (equation [2.16]), is finite throughout the turn-around.

3.4 Sensitivity of Fluctuation Amplitude to the Time of Transition

It is often argued that the amplitude of density fluctuation at late times, the quantity $\delta\rho/\rho|_{\text{asymp}}$ used in equation (2.45) for instance, is unaffected by the details of the transition from the inflationary to the radiation-dominated stage. We demonstrate in this section the possibility that the detail of the transition process does matter. For this purpose we use an instantaneous-reheating-approximation. This is not meant to be realistic but serves the purpose of illustrating the possibilities. The behavior of $\delta\rho/\rho|_{\text{asymp}}$ is investigated for two cases: when inflation ends during the slow-roll phase and when it ends very close to the point of turn-around.

Suppose inflation ends at some time t_{tr} during the slow-roll phase. It is argued in Section 3.2 that in the plateau/slow-roll phase, ζ quickly approaches a constant value ζ_{sl} (equation [3.4]) after Hubble-radius-crossing. Therefore, right before the transition

$$\zeta(t_{\text{tr}}) = \zeta_{\text{sl}} \tag{3.11}$$

According to equation (3.2), $\dot{\zeta} = K_2/(a^3\gamma)$. Hence $K_2 = \dot{\zeta}a^3\gamma\big|_{t_{\text{cr}}}$ where t_{cr} is the time of Hubble-radius-crossing. Φ is related to the derivative of ζ through equation (2.29).

Therefore right before the transition

$$\Phi(t_{\rm tr}) = -\left.\dot{\zeta} a^3 \gamma\right|_{t_{\rm cr}} \left.\frac{H}{ak^2}\right|_{t_{\rm tr}} \tag{3.12}$$

Putting the two expressions above into equations (2.43b) and (2.43c), it can be shown that $F \sim \zeta_{\rm sl}$ and G is negligible to excellent approximation as long as $a(t_{\rm cr}) \ll a(t_{\rm tr})$, that the universe expands by significant amount from Hubble-radius-crossing to the transition. It is also important that γ, although small, does not vanish in the slow-roll phase.

Putting $F - G \sim \zeta_{\rm sl}$ into equation (2.45) and making use of equation (2.27) to relate ζ and $\delta\varphi_{GI}$, the following is obtained:

$$\left.\frac{\delta\rho}{\rho}\right|_{\rm asymp} = 4H\frac{\delta\varphi_{GI}}{\dot{\varphi}_O}\,. \tag{3.13}$$

The quantity on the right hand side can be evaluated at any time after Hubble-radius-crossing as it rapidly approaches a constant value. This gives the amplitude of the acoustic oscillation after the fluctuation enters the Hubble radius in the radiation-dominated stage. This is the formula that has become part of the standard lore of the inflationary theory. A fluctuation that enters the Hubble radius in the matter dominated stage obeys a similar relation with a different numerical constant on the right hand side.

The key to the derivation of the above result is that ζ remains almost constant in the slow-roll phase. ζ is not, strictly speaking, conserved through the transition (equation [2.41]). However, its jump at the transition is small if the inflationary stage ends during the plateau/slow-roll phase. This is because wavelengths of cosmological interests are much longer than the Hubble-radius at reheating and γ, although small, is non-vanishing in the slow-roll phase. That is why we can use $F \sim \zeta_{\rm sl}$ and $G \sim 0$. The constancy of ζ from the exit to the entry of the Hubble radius is a very good approximation for wavelengths of cosmological interests, if inflation ends during the slow-roll phase.

The same is not true if the inflationary stage ends close to the point of turn-

around. It is shown in the last section that ζ diverges close to it. We denote this point by $t = 0$. For sufficiently small t,

$$\zeta \sim \zeta^{II}/t \quad , \quad \Phi \sim \Phi^{II} , \tag{3.14}$$

according to equations (3.9a) and (3.9b). The constant ζ^{II} can be related to K_2 in equation (3.3) by expanding the integrand in powers of t:

$$a^3\gamma = -\left.\frac{a^3}{2H^2}\frac{d^3H}{dt^3}\right|_{t=0} t^2 + ... \tag{3.15}$$

Performing the integration in equation (3.3) using the above expansion, the dominant contribution to ζ in the small t limit is obtained. Comparing it with equation (3.14)

$$\zeta^{II} = \left.\frac{2H^2}{a^3}\left(\frac{d^3H}{dt^3}\right)^{-1}\right|_{t=0} K_2 \tag{3.16}$$

$\dot{\zeta}$ can be calculated from the same equation (3.3). Φ is related to $\dot{\zeta}$ by equation (2.29). Hence

$$\Phi^{II} = -\left.\frac{H}{ak^2}\right|_{t=0} K_2, \tag{3.17}$$

where equation (2.8) has been used.

K_2 is simply equal to $\left.\dot{\zeta}a^3\gamma\right|_{t_{cr}}$ (equation [3.3]). Substituting equation (3.14), together with the expressions obtained above, into equation (2.43), one can obtain

$$F = \left[\dot{\zeta}(t_{cr})\gamma(t_{cr})H^3(0)\left(\frac{d^3H(0)}{dt^3}\right)^{-1}\right]\left[\frac{a^3(t_{cr})}{a^3(0)}\right]\left[\frac{2}{H(0)t} + \frac{2}{3(H(0)t)^2}\right] - G \tag{3.18a}$$

$$G = -\left[\frac{\dot{\zeta}(t_{cr})\gamma(t_{cr})}{6H(0)}\right]\left[\frac{a^3(t_{cr})}{a^3(0)}\right] \tag{3.18b}$$

In obtaining F, we have used $\gamma = 2$ after the transition and $\gamma \sim -t^2(2H(0)^2)^{-1}d^3H(0)/dt^3$ close to the point of turn-around, but before the transition. We have also equated a and H at the time of transition ($t = t_{tr}$) with their respective values at the point of turn-around ($t = 0$), assuming the transition takes place very close to it. This is good enough for our purpose because we are interested in deriving the most diver-

gent contributions in the small t limit (t is now the difference between the times of transition and turn-around).

G is exponentially suppressed because of the factor of $a^3(t_{\rm cr})/a^3(0)$. So is F unless $H(0)t$ is extremely close to zero, which means the transition takes place very near to the point of turn-around.

Let us consider the different contributions to F. $\dot{\zeta}$ can be expressed in terms of the derivative of $H\delta\varphi_{GI}/\dot{\varphi}$ through equation (2.27). Using equations (2.31) and (2.8), it can be shown that $d^3H/dt^3 = -(\ddot{\varphi}_0)^2$ at the point of turn-around. Putting the expression for F into equation (2.45) and ignoring G, one obtains

$$\left.\frac{\delta\rho}{\rho}\right|_{\rm asymp} = 4\left[\frac{a^3(t_{\rm cr})}{a^3(0)}\right]\left[\frac{\dot{\varphi}_0^2}{H^2}\frac{d}{dt}\left(\frac{H\delta\varphi_{GI}}{\dot{\varphi}_0}\right)\right]\Bigg|_{t_{\rm cr}}\left[\frac{-H^3}{\ddot{\varphi}_0^2}\right]\Bigg|_{t=0}\left[\frac{1}{3H^2(0)t^2}\right]. \qquad (3.19)$$

$t_{\rm cr}$ is the time when $k^2/(2a^2H^2) = 1$ (equation [3.2]). $t = 0$ is the time when the inflaton turns around. t is the difference between the time when inflation ends and the time of turn-around. Note that we have kept only the most divergent term: $1/t^2$.

Note that equation (3.19) was derived by setting the integration constant K_2 equal to $\dot{\zeta}a^3\gamma\big|_{t_{\rm cr}}$ (see the paragraph preceding equation [3.18a]). This follows from taking the derivative of equation (3.2) at time $t_{\rm cr}$ (when $k^2 = (a\sqrt{\gamma})''/a\sqrt{\gamma}$). Strictly speaking, equation (3.2) holds only at some time later than $t_{\rm cr}$ because it is an approximation in the regime $k^2 \ll (a\sqrt{\gamma})''/a\sqrt{\gamma}$. Hence, the first two brackets on the right hand side of equation (3.19) should really be evaluated not exactly at the time when $k^2 = (a\sqrt{\gamma})''/a\sqrt{\gamma}$ but some time later when equation (3.2) is a good approximation. Exactly what time one picks is unimportant because as long as equation (3.2) holds, the product of the first two brackets on the right hand side of equation (3.19) is independent of the time at which it is evaluated.

There are a few note-worthy features of equation (3.19). First, the expression is negligible unless t is fine-tuned to be very close to the time of turn-around, because of the exponential suppression due to $a^3(t_{\rm cr})/a^(0)$. This does not mean the density fluctuation vanishes if the t is not close enough to zero: we have dropped terms that are higher order in t in equation (3.14). It can be shown that some of these terms do

not have the exponential suppression factor $a^3(t_{\rm cr})/a^3(0)$.

Second, the amplitude is dependent upon, not $H\delta\varphi_{GI}/\dot{\varphi}_0$ as in the standard formula, but its derivative. In addition, the expression above is also dependent on $\ddot{\varphi}_0$ at the end of inflation. Hence the amplitude of density fluctuations is sensitive to the slope of the potential where the transition takes place.

3.5 Conclusion

Equations (3.13) and (3.19) contain the main results of the first half of this thesis.

If inflation ends during the slow-roll phase, we obtain equation (3.13) as the expression for the amplitude of density fluctuations when they enter the Hubble radius in the radiation-dominated era. It agrees with the standard result. ζ, which obeys equation (2.19), is, to excellent approximation, a conserved quantity throughout reheating. The reason is because the decaying mode for ζ (the mode associated with the constant K_2 in equation [3.2]) is redshifted away by enormous amount of inflation, after Hubble-radius-exit.

However, if inflation ends very close to the point of turn-around, the same mode that decays in the slow-roll phase can actually become dominant. This gives rise to equation (3.19). The amplitude of density fluctuations diverges if inflation ends exactly at the point of turn-around. But since this effect is associated with the "decaying" mode in the slow-roll phase, the divergence is exponentially suppressed. Hence, the divergence dominates only if inflation ends almost exactly at the time of turn-around.

The divergence occurs to fluctuations on all scales. We have therefore demonstrated an exception to the conventional wisdom, that the transition process cannot affect large scale density fluctuations, which have wavelengths much larger than the Hubble radius during reheating.

The break down of the standard lore occurs because ζ or other analogous quantities, which is often argued to remain constant for super-Hubble-radius wave modes, does change close to the point of turn-around. In fact, it diverges. It should be

emphasized, however, that this does not imply any physical divergence unless inflation ends right at the moment of turn-around. We have shown in Section 3.3 that Φ, which is proportional to a gauge invariant density fluctuation, remains finite throughout the turn-around. It is the divergence of ζ at that particular moment, together with the role it plays in the junction conditions, that causes the physical divergence of density fluctuations at second Hubble-radius crossing, if one forces inflation to end precisely at the moment of turn-around.

The divergence is obtained in a simple model of instantaneous transition. In a more realistic calculation, one expects a few changes. First, the great increase in the amplitude of density fluctuations at second Hubble-radius-crossing occurs only if one artificially ends inflation almost exactly at the point of turn-around. In a more realistic calculation, it is quite possible that this effect will not be dominant. Second, the divergence occurs because the sum of pressure and energy, in equation (2.19), vanish when the scalar field momentarily stops at the points of turn-around. If a small amount of radiation is allowed to be present, the denominator in the expression for ζ cannot vanish any more and the divergence is suppressed.

Nonetheless, it remains to be shown whether the reheating process can affect large scale density fluctuations, when the relevant physics of the inflaton-decay is taken into account properly. The very existence of the divergence at the point of turn-around is an interesting indication of what can happen.

Chapter 4

Local Approximations: Formulation

4.1 Introduction

The complexity of nonlinear gravitational instability challenges our understanding of the universe. Even though the law of gravity between two bodies is very simple in the non-relativistic limit, the long-range interactions among exceedingly many bodies leads to behavior that defies simple analysis beyond the linear regime. Computer simulation with N-body methods provides a comprehensive approach to this problem, but it suffers from finite dynamic range and computational expense. Even more importantly, simulations do not increase our understanding of dynamics without guidance from analytical approaches.

In Chapters 4 and 5, we explore a class of what we call local approximations for the nonlinear dynamics of self-gravitating cold matter. By local we mean that the density, velocity gradient, and gravity gradient for each mass element behave as if the element evolves independently of all the others once the initial conditions are specified. This might sound quite implausible. After all, mass elements do influence each other through gravity. However, as we will demonstrate, the celebrated Zel'dovich (1970) approximation (henceforth ZA) can be viewed as exactly an approximation of this sort. (For readers who are familiar with the ZA, it is probably obvious that the ZA is local. Our aim is to cast the ZA in a form useful for generalization to other local approximations.)

In the past several years, there have been various attempts to improve upon the ZA. A notable recent example is the modified Zel'dovich approximation (MZA) proposed by Reisenegger & Miralda-Escudé (1995). The approximation is exact for the gravitational collapse of a mass element with spherical, cylindrical or planar symmetry, just like a new approximation that we propose in this chapter. Unfortunately, the MZA suffers from unphysical singularities for a certain class of otherwise acceptable initial conditions (Reisenegger & Miralda-Escudé 1995).

Other attempts to improve upon the ZA include the adhesion approximation (Kofman et al. 1990), the frozen flow approximation (Matarrese et al. 1992), the frozen potential approximation (Brainerd et al. 1993; Bagla & Padmanabhan 1994), the truncated Zel'dovich approximation (Coles et al. 1993), the smoothed potential approximation (Melott et al. 1996) and higher-order Lagrangian perturbation theory (Melott et al. 1995) (note that the ZA can also be regarded as the first-order solution in Lagrangian perturbation theory). Most of them are attempts to deal with the evolution of high density regions after trajectories cross, when the ZA ceases to be adequate. However, this is a difficult problem. Aside from the spherical model (Peebles 1980) and its cousins, there still exists little in the way of approximation methods for post-collapse evolution.

In this thesis , we will not try to tackle the problem of trajectory crossing or the subsequent nonlinear evolution. Instead we ask whether one can improve upon the ZA even before orbits cross by seeking generalizations of the ZA within the framework of local approximations. In simple terms, a local approximation is one in which the evolution of each mass element is described by a set of ordinary differential equations in time in which there is no coupling to other mass elements, aside from those implied by the initial conditions. For instance, as we will explain more fully later, the evolution of a given mass element under the ZA is completely determined once the initial expansion, vorticity, shear and density at this mass element are specified. (The first three quantities correspond to the trace, antisymmetric part and traceless symmetric part of the velocity gradient tensor.) The evolution of other mass elements have no effect on the evolution of these quantities at this mass element. In other words,

under the ZA, all the information about other mass elements is encoded in the initial conditions. Once these are specified, each mass element goes for its own "free ride"!

We shall seek generalizations of the ZA by first systematically writing down a set of Lagrangian evolution equations for the velocity and gravity gradient for a given fluid element. We discuss two local approximations based on ignoring certain terms in the evolution equation for the Newtonian tidal tensor. One of them was introduced by Bertschinger & Jain (1994). The history associated with it has been discussed in Chapter 1 and is not repeated here.

The second local approximation based on the tidal evolution equation is entirely new. It is based on dropping several more terms in addition to the Weyl tensor term. We will show why this is a better approximation compared to the one proposed by Bertschinger & Jain (1994). In fact, in tests this new approximation performs even better than the ZA, both in cases where exact solutions are known and where numerical solutions are calculated. In this thesis, we concentrate on a comparison of the three local approximations for ellipsoids, with and without symmetries.

To understand the main ideas underlying these local approximation methods, and how they differ from other approaches, it is useful to draw an analogy with gravitational lensing. Our use of Lagrangian fluid equations is akin to solving the optical scalar equations (Sachs 1961), whereby one follows the two-dimensional cross-section of a congruence of light rays propagating through space. Our approach is similar, with light rays replaced by cold dust, and with the two-dimensional cross-section replaced by the three-dimensional volume of a mass element. In fact, both approaches follow from the pioneering work in general relativity by Ehlers (1961) and Kundt & Trümper (1961). The first application of these methods to matter was by Hawking (1966), who pioneered the covariant fluid approach to cosmological perturbation theory. The formalism was championed by Ellis (1971) and eventually was applied to the formation of large scale structure (see Bertschinger 1995 and the references cited previously). As in the case of gravitational lensing, this approach can tell how a given (mass) element evolves but does not give its trajectory. The optical scalar equations do not replace the gravitational lens equation, they supplement it. Like-

wise, the local methods can supplement N-body simulations or other approximations such as Lagrangian perturbation theory, by providing accurate ways to follow the deformation of mass elements as they evolve under gravity.

The organization of this chapter is as follows. In Section 4.2, we show how the ZA is a local approximation. Section 4.3 presents two additional local approximations based on dropping terms from the tidal evolution equation, and shows under what circumstances these approximations are exact. In the next chapter, we describe tests to compare the three approximations. Chapters 4 and 5 are largely based on a preprint (Hui & Bertschinger 1996).

4.2 On the Zel'dovich Approximation

In this section we review the Zel'dovich approximation starting from the Eulerian fluid equations in comoving coordinates. We then show that it can be regarded as a local approximation.

The cosmological fluid equations for cold dust in a perturbed Robertson-Walker universe with expansion scale factor $a(\tau)$ are (Bertschinger 1995):

$$\frac{\partial \delta}{\partial \tau} + \nabla_i \left[(1+\delta) v^i \right] = 0 \ , \tag{4.1}$$

$$\frac{\partial v^i}{\partial \tau} + v^j \nabla_j v^i = -\frac{a'}{a} v^i - \nabla^i \phi \ , \tag{4.2}$$

$$\nabla^2 \phi = 4\pi G a^2 \rho_0 \delta \ . \tag{4.3}$$

The mass density is $\rho = \rho_0(\tau)(1+\delta)$ and $\vec{v} = d\vec{x}/d\tau$ is the proper peculiar velocity where $\vec{x}$ is the comoving spatial position and τ is the conformal time (hence, $d\tau = dt/a$ where t is the proper time). We are neglecting spatial curvature so that we can use Cartesian coordinates where $\nabla^i = \nabla_i = \partial/\partial x^i$ for the ith spatial coordinate. The prime denotes differentiation with respect to conformal time.

The trajectory of a fluid element is $x^i(\vec{q}, \tau)$ where $\vec{q}$ is a Lagrangian coordinate

labeling the element, conventionally chosen to be the initial position:

$$x^i(\vec{q},\tau) = q^i + \psi^i(\vec{q},\tau) \ . \tag{4.4}$$

Now we introduce the Lagrangian time derivative $d/d\tau \equiv \partial/\partial\tau + v^j\nabla_j$. This time derivative commutes with $\partial/\partial q^i$. Using $v^i = d\psi^i/d\tau$, we can rewrite equation (4.2) as

$$\frac{d^2}{d\tau^2}\psi^i + \frac{a'}{a}\frac{d}{d\tau}\psi^i - 4\pi G a^2\rho_0\psi^i = -\nabla^i\phi - 4\pi G a^2\rho_0\psi^i \ . \tag{4.5}$$

Each term on the left-hand side is first order in ψ^i. The right-hand side can be estimated from the Poisson equation (4.3), but first we need the mass density. It follows in the Lagrangian approach by noting that ρd^3x is conserved along a fluid streamline provided d^3x is computed from the mapping $\vec{q} \to \vec{x}$. If there are no displacements, $\vec{q} = \vec{x}$ and $\rho = \rho_0$. The volume element follows from the Jacobian determinant, leading to

$$\rho(\vec{q},\tau) = \rho_0 \left|\frac{\partial x^i}{\partial q^j}\right|^{-1} \ . \tag{4.6}$$

For small displacements the Jacobian may be expanded in a power series; the first-order term gives $\rho = \rho_0(1-\partial\psi^i/\partial q^i)+O(\psi^2)$. Now note that $\partial\psi^i/\partial x^i = (\partial\psi^i/\partial q^j)(\partial q^j/\partial x^i) = \partial\psi^i/\partial q^i + O(\psi^2)$. Therefore, using equation (4.3), we see that the divergence of the right-hand side of equation (4.5) vanishes to first order in ψ^i. If ψ^i is longitudinal (i.e., has vanishing curl), then the right-hand side itself vanishes to first order. Displacements that grow by gravity are necessarily longitudinal in linear theory. The ZA consists of setting to zero the right-hand side of equation (4.5). (It can be generalized to allow for a transverse displacement; see Buchert 1993 and Barrow & Saich 1993 .) Under the ZA, the evolution of displacement thus obtained is used in equation (4.6) to get the density field. The ZA is equivalent to first order Lagrangian perturbation theory for the trajectories $\vec{x}\,(\vec{q},\tau)$.

With vanishing right-hand side, equation (4.5) is identical to the linear perturbation evolution equation for δ (a fact that becomes obvious when one notes $\delta = -\partial\psi^i/\partial q^i$ and $d/d\tau = \partial/\partial\tau$ to first order in ψ). This second-order ordinary

differential equation in time has two independent solutions that we write $D_{\pm}(\tau)$ (Peebles 1980). Taking the growing solution and requiring ψ^i to be longitudinal, we get the solution

$$\psi^i(\vec{q},\tau) = D_+(\tau)\frac{\partial\varphi(\vec{q})}{\partial q^i} \tag{4.7}$$

where $\varphi(\vec{q})$ is a displacement potential which is fixed by initial conditions.

Next we will show that equations (4.4) and (4.7) imply that the ZA displacement field is longitudinal in x-space (the irrotational initial conditions already imply it is irrotational in q-space), a first step needed before we show that the ZA is a local approximation. We have

$$\left(\vec{\nabla}\times\vec{\psi}\right)_i = D_+(\tau)\epsilon_{ijk}\frac{\partial}{\partial x^j}\left(\frac{\partial\varphi}{\partial q^k}\right) = D_+\epsilon_{ijk}\left(\frac{\partial q^l}{\partial x^j}\right)\left(\frac{\partial^2\varphi}{\partial q^k\,\partial q^l}\right) ,$$

where ϵ_{ijk} is the usual antisymmetric Levi-Civita symbol. Now, note that the Jacobian matrix defined by the transformation of equations (4.4) and (4.7), $\partial x^j/\partial q^l = \delta_{jl} + D_+\partial^2\varphi/\partial q^j\,\partial q^l$, is real and symmetric. By a theorem of linear algebra its inverse, $\partial q^l/\partial x^j$, is also symmetric. So is $\partial^2\varphi/\partial q^k\,\partial q^l$ and, because they commute, so is their product. Thus, in the equation for $\vec{\nabla}\times\vec{\psi}$ above, ϵ_{ijk} is contracted with a matrix that is symmetric in j and k, yielding $\vec{\nabla}\times\vec{\psi} = 0$ (Zel'dovich & Novikov 1983).

The implication of this result is that $\vec{\psi}$ is longitudinal in $\vec{x}$-space as well as in $\vec{q}$-space. The same conclusions hold for the velocity field $\vec{v}$, since it differs from $\vec{\psi}$ by only a time-varying factor D'_+/D_+. As a result, under the Zel'dovich approximation we can write

$$\psi^i(\vec{q}(\vec{x},\tau),\tau) = D_+(\tau)\frac{\partial\Phi(\vec{x},\tau)}{\partial x^i} \quad \text{and} \quad v^i(\vec{x},\tau) = D'_+(\tau)\frac{\partial\Phi(\vec{x},\tau)}{\partial x^i} \; ; \quad \frac{d\Phi(\vec{x},\tau)}{d\tau} = 0 \; . \tag{4.8}$$

The last equation follows from the fact that $\partial\Phi/\partial x^i = \partial\varphi/\partial q^i$ (cf. eq. 4.7). Recall that under the ZA the right-hand side of equation (4.5) vanishes. Using equation (4.8), we then get

$$\vec{v} = -D'_+\left(4\pi G a^2\rho_0 D_+\right)^{-1}\vec{\nabla}\phi = -\frac{2a'\,f}{3\Omega_0 H_0^2}\vec{\nabla}\phi \; , \tag{4.9}$$

where $f \equiv d\ln D_+/d\ln a$. Thus, in the Zel'dovich approximation, the velocity field is always (not just to first order in $\vec{\psi}$) proportional to the gravity field (Kofman 1991). It is clear geometrically that this result must be correct for planar, cylindrical, or spherical flow for growing mode initial conditions. For plane-parallel flows, but not otherwise, the coefficient of proportionality of the ZA is also correct, so that the ZA is exact (until orbit-crossing) in one dimension.

We are now going to present the ZA from another point of view. Similar work has been done by Kofman & Pogosyan (1995). Our aim is to motivate how one might improve the ZA by generalizing it to a broader class of local approximations. It will become clear shortly exactly what we mean by local approximations.

Let us first give a brief summary of the Lagrangian fluid equations (Bertschinger & Jain 1994). First of all, the gradient of the fluid velocity field is decomposed into its trace, traceless symmetric and antisymmetric parts, which are the expansion θ, shear σ_{ij} and vorticity ω_{ij} respectively:

$$\nabla_i v_j = \frac{1}{3}\,\theta\,\delta_{ij} + \sigma_{ij} + \omega_{ij} \quad \sigma_{ij} = \sigma_{ji}\ , \quad \omega_{ij} = \epsilon_{ijk}\,\omega^k = -\omega_{ji}\ , \tag{4.10}$$

where $2\vec{\omega} = \vec{\nabla} \times \vec{v}$. Then, converting time derivatives from Eulerian to Lagrangian, equation (4.1) becomes

$$\frac{d\delta}{d\tau} + (1+\delta)\,\theta = 0 \tag{4.11}$$

Taking the trace of equation (4.2) and using equations (4.3) and (4.10), one obtains the Raychaudhuri equation:

$$\frac{d\theta}{d\tau} + \frac{a'}{a}\,\theta + \frac{1}{3}\,\theta^2 + \sigma^{ij}\sigma_{ij} - 2\omega^2 = -4\pi G a^2 \rho_0 \delta\ , \tag{4.12}$$

where $\omega^2 \equiv \omega^i\omega_i$. Similarly, taking the antisymmetric and traceless symmetric parts of equation (4.2) gives respectively

$$\frac{d\omega^i}{d\tau} + \frac{a'}{a}\,\omega^i + \frac{2}{3}\,\theta\,\omega^i - \sigma^i{}_j\,\omega^j = 0 \tag{4.13}$$

and

$$\frac{d\sigma_{ij}}{d\tau} + \frac{a'}{a}\,\sigma_{ij} + \frac{2}{3}\,\theta\,\sigma_{ij} + \sigma_{ik}\sigma^{k}{}_{j} + \omega_i\omega_j - \frac{1}{3}\,\delta_{ij}\left(\sigma^{kl}\sigma_{kl} + \omega^2\right) = -E_{ij}\ , \qquad (4.14)$$

where $E_{ij} \equiv \nabla_i\nabla_j\phi - (1/3)\,\delta_{ij}\,\nabla^2\phi$ is the gravitational tidal field.

In keeping with the spirit of Lagrangian fluid dynamics, we would like an evolution equation for E_{ij}. From equations (4.1) and (4.3), Bertschinger & Hamilton (1994) derived

$$\frac{dE_{ij}}{d\tau} + \frac{a'}{a}\,E_{ij} - \nabla_k\,\epsilon^{kl}{}_{(i}H_{j)l} + \theta E_{ij} + \delta_{ij}\,\sigma^{kl}E_{kl} - 3\sigma^{k}{}_{(i}E_{j)k} - \omega^{k}{}_{(i}E_{j)k} = -4\pi G a^2\rho\,\sigma_{ij}\ . \qquad (4.15)$$

Parentheses around a pair of subscripts indicates symmetrization, e.g., $\sigma^{k}{}_{(i}E_{j)k} = (\sigma^{k}{}_{i}E_{jk} + \sigma^{k}{}_{j}E_{ik})/2$. The new quantity H_{ij} is the Newtonian limit of the magnetic part of the Weyl tensor in the fluid frame. The definition and discussion of this term will be deferred until the next section.

Equations (4.11) to (4.15) form a hierarchy of Lagrangian fluid equations. It is an incomplete set because we have not stated the evolution equation for H_{ij}. In order to arrive at a local set, we must eliminate the gradient term in equation (4.15), either by finding an approximation for $-\nabla_k\,\epsilon^{kl}{}_{(i}H_{j)l}$ or by truncating the hierarchy in a way that eliminates our need to determine it.

The ZA eliminates the need to calculate H_{ij} by approximating the evolution of the gravity field — equation (4.9) relates $\vec{\nabla}\phi$ to $\vec{v}$. As a result, the tidal tensor in the ZA follows from the shear:

$$E_{ij} = -\frac{4\pi G a\rho_0}{Hf}\sigma_{ij} = -\frac{3\Omega_0 H_0^2}{2a'f}\sigma_{ij}\ . \qquad (4.16)$$

Furthermore, the divergence of the gravity field is given in the ZA by the velocity expansion scalar θ instead of the density fluctuation. Thus, the ZA is equivalent to solving the local evolution equations

$$\frac{d\theta}{d\tau} + \frac{a'}{a}\,\theta + \frac{1}{3}\,\theta^2 + \sigma^{ij}\sigma_{ij} - 2\omega^2 = \frac{4\pi G a\rho_0}{Hf}\theta\ , \qquad (4.17)$$

$$\frac{d\sigma_{ij}}{d\tau} + \frac{a'}{a}\,\sigma_{ij} + \frac{2}{3}\,\theta\,\sigma_{ij} + \sigma_{ik}\sigma^{k}{}_{j} + \omega_i\omega_j - \frac{1}{3}\,\delta_{ij}\left(\sigma^{kl}\sigma_{kl} + \omega^2\right) = \frac{4\pi G a \rho_0}{Hf}\sigma_{ij}\ . \quad (4.18)$$

Together with equations (4.11) and (4.13), these give a closed set of equations for the evolution of quantities for a single mass element with no spatial gradients. This is what we mean by locality. Note that we have assumed the irrotational flow initial condition and so $\omega_{ij} = 0$ from equation (4.13) at all times before trajectories intersect. Equation (4.18) can be written also as an evolution equation for E_{ij} by making use of equation (4.16) (Kofman & Pogosyan 1995). But it is clear that in terms of obtaining a closed set of local equations, it is sufficient to stop at the level of the shear equation (4.18).

Hence, we have shown that the ZA is a local approximation based on truncating the set of Lagrangian fluid equations at the shear evolution equation by setting E_{ij} proportional to σ_{ij} and by approximating the gravitational source term in the Raychaudhuri equation. It is then very natural to ask whether we can go further, by using the exact Raychaudhuri equation and by truncating the system of equations at the tidal evolution equation with a different approximation from the ZA.

There is a simple argument for why we should expect to be able to improve on the ZA. It is well known that the ZA gives incorrect results for spherical infall. For spherical infall, the velocity and gravity fields are isotropic around a point, so that $\sigma_{ij} = E_{ij} = 0$ at that point. Yet, the ZA overestimates the collapse time for a uniform spherical tophat. The reason for this is that the ZA does not obey the Poisson equation, so the right-hand side of equation (4.17) is not exact. We can at least correct this term. We have tested this approximation — using equation (4.12) in place of equation (4.17), and using equation (4.18) for the shear evolution — and found that it works poorly aside from spherical flow. Thus, we seek improved approximations based on a more accurate treatment of the tidal tensor.

4.3 Two Local Approximations Based on the Tidal Evolution Equation

As remarked in the last section, the hierarchy of Lagrangian fluid equations can be truncated at the tidal evolution equation, provided that we approximate, or eliminate, the H_{ij} term (and possibly other terms also). If possible, we would like to find local approximations that retain the successes of the Zel'dovich approximation. These include giving the correct results in linear perturbation theory and giving the exact solution for plane-parallel flows. Ideally, we would also like to improve on the Zel'dovich approximation by giving exact results for spherical and/or cylindrical flows. We use these criteria in seeking improved approximations.

Let's look at the magnetic part of the Weyl tensor more closely. The definition is given in Bertschinger & Hamilton (1994):

$$H_{ij} \equiv -\frac{1}{2}\nabla_{(i}H_{j)} - 2\,v_k\,\epsilon^{kl}{}_{(i}E_{j)l} \tag{4.19}$$

where H_i satisfies:

$$\vec{\nabla}\times\vec{H} = -16\pi G a^2\,\vec{f}_\perp\ , \quad \vec{\nabla}\cdot\vec{H} = 0\ . \tag{4.20}$$

Here $\vec{f}_\perp$ is the transverse part of the mass current, defined as follows:

$$\vec{f}_\perp \equiv f - f_\parallel = \rho\vec{v} - f_\parallel\ , \quad \vec{f}_\parallel = -\frac{1}{4\pi G a^3}\vec{\nabla}\left(\frac{\partial\, a\phi}{\partial\tau}\right). \tag{4.21}$$

Using these definitions, we can rewrite equation (4.15) as follows:

$$\frac{dE_{ij}}{d\tau} + \frac{a'}{a}E_{ij} + M_{ij} = -4\pi G a^2\rho\sigma_{ij}\ , \tag{4.22}$$

where

$$\begin{aligned} M_{ij} &\equiv -\nabla_k\,\epsilon^{kl}{}_{(i}H_{j)l} + \theta E_{ij} + \delta_{ij}\,\sigma^{kl}E_{kl} - 3\sigma^k{}_{(i}E_{j)k} - \omega^k{}_{(i}E_{j)k} \\ &= -4\pi G a^2\rho\nabla_{(i}v_{j)} - \frac{1}{a}\frac{d}{d\tau}(\nabla_i\nabla_j a\phi) \end{aligned}$$

$$= \quad -4\pi G a^2 \nabla_{(i} f_{\perp j)} - v_k \nabla^k \nabla_i \nabla_j \phi + v_{(i} \nabla_{j)} \nabla^2 \phi \ . \tag{4.23}$$

Let us first consider plane-parallel flows, for which the ZA is exact. The velocity and gravity gradient tensors may be written

$$\nabla_i v_j = \theta \, \mathrm{diag}(0,0,1) \ , \quad \nabla_i \nabla_j \phi = \nabla^2 \phi \, \mathrm{diag}(0,0,1) \ , \tag{4.24}$$

where diag() denotes the elements of the diagonalized tensor. Evaluating M_{ij} using equations (4.23), we find that the curl H_{jl} term as well as the sum of terms proportional to the tidal tensor vanish identically. The individual tidal terms do not vanish. This result suggests two different closure schemes for the tidal evolution equation (4.22). The first one is to discard $\nabla_k \, \epsilon^{kl}{}_{(i} H_{j)l}$. The second is to discard the complete tensor M_{ij}. If some of the tidal terms of M_{ij} were retained, the resulting approximation would not be exact for one-dimensional flows, hence would not improve on the Zel'dovich approximation.

The first choice, setting $H_{ij} = 0$ in equation (4.15), was proposed by Bertschinger & Jain (1994):

$$\frac{dE_{ij}}{d\tau} + \frac{a'}{a} E_{ij} + \theta E_{ij} + \delta_{ij} \, \sigma^{kl} E_{kl} - 3\sigma^k{}_{(i} E_{j)k} - \omega^k{}_{(i} E_{j)k} = -4\pi G a^2 \rho \, \sigma_{ij} \ . \tag{4.25}$$

We shall call this the non-magnetic approximation (NMA). Combined with equations (4.11)–(4.14), it provides a closed set of local evolution equations. The NMA was inspired, in part, by the remark of Ellis (1971) that the magnetic part of the Weyl tensor has no Newtonian counterpart. However, it leads to unusual behavior, implying that cold dust fluid elements generically collapse to spindles (Bertschinger & Jain 1994). Also, Bertschinger & Hamilton (1994) were able to derive equation (4.15) with H_{ij} defined using equations (4.19)–(4.21) from Newton's laws in an expanding universe, as well as constraint and evolution equations for H_{ij} itself (the latter using post-Newtonian corrections), from which we now know that H_{ij} is not identically zero in the Newtonian limit, aside from some special cases of high symmetry.

Thus, we are motivated to try the second approximation, setting $M_{ij} = 0$ in

equation (4.22):

$$\frac{dE_{ij}}{d\tau} + \frac{a'}{a}E_{ij} = -4\pi G a^2 \rho \sigma_{ij} \ . \tag{4.26}$$

Equation (4.26) and equations (4.11)–(4.14) form our new set of closed local equations. We shall call this the Local Tidal Approximation (LTA) to distinguish it from equation (4.25), the non-magnetic approximation.

The LTA is, in a sense, a close cousin of the NMA. Recall the definition of H_{ij} in equation (4.19). Substitute this into the definition of M_{ij} in equation (4.23). All terms involving the product of E_{ij} and the velocity gradient tensor cancel. The result is

$$M_{ij} = -\nabla_k \epsilon^{kl}{}_{(i} H'_{j)l} + v_m \nabla_k \left[\epsilon^{kl}{}_i \epsilon^{mn}{}_{(j} E_{l)n} + \epsilon^{kl}{}_j \epsilon^{mn}{}_{(i} E_{l)n} \right] \ , \tag{4.27}$$

where $H'_{ij} \equiv H_{ij} + 2v_k \, \epsilon^{kl}{}_{(i} E_{j)l}$ is gotten by setting $v_k = 0$ in equation (4.19). While H_{ij} represents the magnetic part of the Weyl tensor in the fluid threading, H'_{ij} represents it in the comoving threading (Bertschinger & Hamilton 1994).

The distinction between H_{ij} and H'_{ij} is most clear using the $1+3$ threading split of spacetime (Bertschinger 1995). The electric and magnetic parts of the Weyl tensor follow by projecting the Weyl tensor and its dual using a timelike unit vector u^μ (threading) onto the spatial hypersurface orthogonal to u^μ. An analogous procedure is used in electromagnetism to obtain the components of the electric and magnetic field from the field strengh tensor. The magnetic part of the Weyl tensor in the fluid threading, H_{ij}, follows from using the fluid 4-velocity, which has components $u^\mu = a^{-1}(1, v^i)$. Using instead the 4-velocity of observers stationary in the given coordinate system (the Poisson gauge in this case; see Bertschinger and Hamilton 1994), $u'^\mu = a^{-1}(1, 0)$, yields the magnetic tensor H'_{ij} in the comoving threading.

Either threading may be used in general relativity. Fluid threading leads to equation (4.15). Comoving threading leads to

$$\frac{dE_{ij}}{d\tau} + \frac{a'}{a} E_{ij} - \nabla_k \, \epsilon^{kl}{}_{(i} H'_{j)l} + v_m \nabla_k \left[\epsilon^{kl}{}_i \epsilon^{mn}{}_{(j} E_{l)n} + \epsilon^{kl}{}_j \epsilon^{mn}{}_{(i} E_{l)n} \right] = -4\pi G a^2 \rho \, \sigma_{ij} \ . \tag{4.28}$$

Equations (4.15) and (4.28) can be shown to be equivalent by using equations (4.23) and (4.27).

Comparing equations (4.26) and (4.28), we see that the LTA makes two approximations: (1) The curl of the magnetic part of the Weyl tensor vanishes in the comoving threading ($\nabla_k \epsilon^{kl}{}_{(i} H'_{j)l} = 0$), and (2) the advective terms on the left hand side of equation (4.28) vanish (the bracketed terms vanish or else v_m vanishes) at the position and moment of interest. By contrast, the NMA assumes that the magnetic part vanishes *in the fluid threading.* At this stage we cannot say which is a better approximation. For that we must compare both with exact solutions.

The LTA, like the ZA, is exact for plane-parallel flows prior to the intersection of orbits. What about spherically and cylindrically symmetric flows, for which the ZA is not exact? For the LTA, we use equations (4.23) to evaluate M_{ij} for flows that are spherically symmetrical around the fluid element under consideration. As long as the gravity gradient is finite at the origin (a condition that holds for any continuous finite-density mass distribution), this restriction implies that all three eigenvalues of $\nabla_i \nabla_j \phi$ are equal, so $E_{ij} = 0$ identically (similarly $\sigma_{ij} = 0$). Equations (4.25) and (4.26) are satisfied trivially. Thus, the LTA is exact for spherical mass elements. So is the NMA.

Next we consider a non-singular fluid element on the symmetry axis of a cylindrically symmetric flow. By this we mean that two eigenvalues of $\nabla_i \nabla_j \phi$ are equal and the third one vanishes and similarly for $\nabla_i v_j$. In this case we have

$$\nabla_i v_j = \frac{1}{2}\,\theta\,\mathrm{diag}(1,1,0)\ , \quad \nabla_i \nabla_j \phi = \frac{1}{2}\,\nabla^2\phi\,\mathrm{diag}(1,1,0)\ . \tag{4.29}$$

Using this, it is easy to show that the sum of tidal terms in the first form of equation (4.23) do not vanish, while, with equations (4.3) and (4.11), the second form for M_{ij} leads to $M_{ij} = 0$. Thus, the LTA is also exact for cylindrical flows, while the NMA is not. Bertschinger & Jain (1994) erred in concluding that the NMA was exact for cylindrical flows even though they were correct in saying that $H_{ij} = 0$ for such an element. Although H_{ij} vanishes, its curl (taking H_{ij} to be defined everywhere in

the fluid threading) does not. However, when expressed in the comoving threading discussed before, the curl of the magnetic part of the Weyl tensor does vanish for a cylindrical flow (hence the LTA is exact in this case).

One can generalize these results to show from the second form of equation (4.23) that $M_{ij} = 0$ for any flow for which $(\nabla^2\phi)^{-1}\nabla_i\nabla_j\phi$ equals $\theta^{-1}\nabla_{(i}v_{j)}$ and is a constant tensor. These conditions are equivalent to saying that the orientation and axis ratios of the gravitational and velocity equipotentials are constant for the mass element under consideration. Thus, the LTA is exact for flows with equipotentials of constant shape. Although this condition does not always hold, it is valid for the growing mode in the linear regime and it includes spherically and cylindrically symmetric flows as well as plane-parallel flows. Moreover, the gravitational potential contours are more nearly spherical than the density contours around a peak, so their shape would be expected to change relatively slowly with time, suggesting that the LTA may be a good approximation in general.

In the linear regime, the LTA, NMA, and ZA all agree. It is already clear that they must differ in second-order perturbation theory; in the Appendix of the next chapter we present the calculation of $\nabla_k \epsilon^{kl}{}_{(i}H_{j)l}$ and M_{ij}. However, it is more important to see how these various approximations behave as collapse is approached. We know already that generic initial conditions lead to collapse along one dimension (pancake) with the ZA while the NMA leads to collapse along two dimensions (spindle) (Bertschinger & Jain 1994). What about the LTA? How accurate is the LTA for asymmetrical initial conditions? Before answering these questions we first examine the relative sizes of the terms in equation (4.22) for an overdense homogeneous ellipsoid in an expanding universe.

Chapter 5

Local Approximations: Tests

5.1 Introduction

Three different local approximations were introduced in the last chapter. Some of them were shown to be exact in the presence of certain symmetries. For instance, the Zeldovich Approximation (ZA) is exact in cases of planar symmetry while the Local Tidal Approximation (LTA) is exact if the gravitational and velocity potentials have spherical, cylindrical or planar symmetry at the mass element of interest.

In this chapter, we compare the three different local approximations for more general initial conditions. In Section 5.2 we consider the motion of a homogeneous ellipsoid, in both cosmological (Friedmann-Robertson-Walker background) and non-cosmological (vacuum) contexts. The Weyl tensor and other relevant terms in the tidal evolution equation are evaluated. In Section 5.3 we discuss how different non-linear approximations predict pancake versus spindle collapse from generic initial conditions, for which we also calculate the collapse times. Conclusions are presented in Section 5.4. In Appendix C we present some relevant results of second-order perturbation theory.

5.2 Collapse of a Homogeneous Ellipsoid

We summarize here the equations of motion for an irrotational homogeneous ellipsoid embedded in an expanding universe. The various interesting quantities in the tidal evolution equation are then calculated for the collapse of a particular ellipsoid.

We consider an irrotational homogeneous ellipsoid with proper axis lengths R_1, R_2, and R_3 embedded in a Friedmann-Robertson-Walker background. The equations of motion are (Icke 1973; White & Silk 1979):

$$\frac{d^2R_i}{dt^2} = -2\pi G R_i \left[\frac{2}{3}\rho_b + \alpha_i(\rho_e - \rho_b)\right] , \tag{5.1}$$

where t is the proper time ($dt = a d\tau$) and α_i is defined by

$$\alpha_i = R_1 R_2 R_3 \int_0^\infty \frac{ds}{(R_i^2 + s)\sqrt{(R_1^2 + s)(R_2^2 + s)(R_3^2 + s)}} . \tag{5.2}$$

Here ρ_e is the total density within the ellipsoid while ρ_b is the density of the expanding universe surrounding the ellipsoid. They are related to the mean and perturbed densities used previously by $\rho_b = \bar{\rho}$ and $\rho_e = \bar{\rho}(1 + \delta)$. We evaluate them from the evolution of the axis lengths and the background expansion scale factor:

$$\rho_e R_1 R_2 R_3 = \rho_{eo} , \quad \rho_b a^3 = \rho_{bo} , \tag{5.3}$$

where ρ_{eo} and ρ_{bo} are constants. Note that $\alpha_1 + \alpha_2 + \alpha_3 = 2$ and we assume $\Omega_0 = 1$. Note also that since equation (5.1) is second order, there are in general two independent modes. We choose growing mode initial conditions. For small a the first order solution or, equivalently, the ZA result, is

$$R_i(t) = a(t) X_i \left(1 - \frac{1}{2}\alpha_{i0}\delta_0 a\right) \tag{5.4}$$

where the X_i's give the initial axis ratios, α_{i0} gives the initial ellipsoid parameter, and δ_0 gives the linear amplitude of the density perturbation. We set $\delta_0 = 1$ for overdense

ellipsoids without loss of generality.

Equation (5.3) implies that the total mass, including the mass inside the ellipsoid as well as outside, is actually not conserved even though the mass inside the ellipsoid is. Hence equation (5.1) can only be an approximation to the true evolution of an initially homogeneous ellipsoid. In general, one expects that such an ellipsoid would cause the density of its immediate surroundings to deviate from the cosmic mean. Tidal fields from this perturbed external material should then induce departure from homogeneity in the ellipsoid. Based on results from an N-body simulation (S.D.M. White, unpublished notes), we assume that it is a good approximation to ignore departures from homogeneity inside and outside the ellipsoid when calculating the evolution of the axis ratios.

It is noteworthy that equation (5.1) is exact if $\rho_b = 0$, i.e., for a homogeneous ellipsoid in a vacuum. Later in this section we will test our approximations using the exact solution in this case.

The peculiar velocity field inside the homogeneous ellipsoid is described by:

$$v_i = \left(\frac{R_i'}{R_i} - \frac{a'}{a} \right) x_i \tag{5.5}$$

and the gravitational potential within the ellipsoid is:

$$\phi = \pi G a^2 (\rho_e - \rho_b) \sum_i \alpha_i x_i^2 \ . \tag{5.6}$$

(The x_i are comoving coordinates and primes denote conformal time derivatives.) Quantities like the expansion, shear, and tidal field can be immediately read off from these expressions:

$$\theta = \sum_i \frac{R_i'}{R_i} - 3\frac{a'}{a} \ , \tag{5.7}$$

$$\sigma_{ij} = \mathrm{diag} \left(\frac{R_i'}{R_i} - \frac{1}{3} \sum_k \frac{R_k'}{R_k} \right) \ , \tag{5.8}$$

$$E_{ij} = 2\pi G a^2 (\rho_e - \rho_b) \, \mathrm{diag} \left(\alpha_i - \frac{2}{3} \right) \ . \tag{5.9}$$

The tensor M_{ij} defined in equation (4.23) is given for the homogeneous ellipsoid by

$$M_{ij} = 2\pi G a^2 \rho_e \,\mathrm{diag}\left[-2\sigma_{ij} + \left(\alpha_i - \frac{2}{3}\right)\theta - \frac{\delta}{1+\delta}\,\alpha_i'\right] \ . \tag{5.10}$$

Using the time evolution of R_i given by equation (5.1), the evolution of the various quantities above can be calculated. In particular, we are interested in the relative magnitude of various terms in the tidal evolution equations (4.15) and (4.22). We integrated equations (5.1) and (5.2) numerically starting from equation (5.4) at $a = 10^{-8}$ with axis ratios $1 : 1.25 : 1.5$. From the axis lengths R_i and their time derivatives, using equations (5.7)–(5.9) we calculated the velocity and gravity gradient terms inside the ellipsoid. From these we then calculated the evolution of $-\nabla_k \epsilon^{kl}{}_{(i} H_{j)l}$, M_{ij}, and $4\pi G a^2 \rho \sigma_{ij}$ inside the ellipsoid. Note that in this test we do not integrate the tidal evolution equation itself; rather, we evaluate the terms in it assuming that the system evolves according to the homogeneous ellipsoid solution. Although, as we noted above, this solution is not exact, we are being self-consistent by evaluating the various tensor quantities using equations (5.7)–(5.10), which assume spatial homogeneity inside the ellipsoid.

Figure 1 shows the results of this calculation. The magnitudes of $-\nabla_k \epsilon^{kl}{}_{(i} H_{j)l}$ and M_{ij} are divided by the magnitude of $4\pi G a^2 \rho \sigma_{ij}$, where by the magnitude of a matrix we mean the square root of the trace of its square. We see that $-\nabla_k \epsilon^{kl}{}_{(i} H_{j)l}$ and M_{ij} are both small compared to $4\pi G a^2 \rho \sigma_{ij}$ at both early and late times, but not intermediate times (near maximum expansion). Interestingly, the magnetic term and M_{ij} have similar magnitude throughout the collapse process.

We can easily understand why $4\pi G a^2 \rho \sigma_{ij}$ is much bigger than both $-\nabla_k \epsilon^{kl}{}_{(i} H_{j)l}$ and M_{ij} at early times using perturbation theory. The shear is first order. The last form of equations (4.23) is the best place to see that M_{ij} is second order: $f_{\perp i}$ is second order because, to first order, $f_i = \bar{\rho} v_i$ is longitudinal (we assume irrotational initial conditions). The other contributions to M_{ij} are obviously second order. From the first form of equation (4.23), we conclude also that $-\nabla_k \epsilon^{kl}{}_{(i} H_{j)l}$ is second order. Expressions for these two tensors in second order perturbation theory are given in

Appendix C. They are both nonzero in general.

The behavior of these quantities close to the moment of pancake collapse can also be estimated analytically. Suppose that the third axis collapses while the other two axes still have finite lengths. The term $4\pi Ga^2\rho\sigma_{ij}$ diverges at the moment of pancake collapse because ρ diverges and so does σ_{ij}, owing to the R_3'/R_3 term in equation (5.8). For the behavior of $-\nabla_k \epsilon^{kl}{}_{(i}H_{j)l}$ and M_{ij} approaching collapse, we need to understand the behavior of the α_i's.

It follows from equation (5.2) that α_1 and α_2 vanish in the limit of vanishing R_3 for finite R_1 and R_2, because the integral is finite while the factor of R_3 in front vanishes. Note also, by definition, the three α_i's always add up to 2. Hence $\alpha_3 = 2$ at collapse. Moreover, it can be verified using equation (5.2) that the α_i''s are finite at the moment of collapse, assuming the R_i''s are finite. It can then be shown using equations (5.7)–(5.10) that the particular combination of α_i's conspires to render both $M_{ij}/4\pi Ga^2\rho$ and $-\nabla_k \epsilon^{kl}{}_{(i}H_{j)l}/4\pi Ga^2\rho$ finite. Hence at the moment of pancake collapse, $-\nabla_k \epsilon^{kl}{}_{(i}H_{j)l}$ and M_{ij} are indeed much smaller than $4\pi Ga^2\rho\sigma_{ij}$.

The fact that $-\nabla_k \epsilon^{kl}{}_{(i}H_{j)l}$ and M_{ij} are both small compared to $4\pi Ga^2\rho\sigma_{ij}$ at early times and at the moment of pancake collapse suggests that the NMA and LTA might both be good approximations. However, Figure 1 shows that these terms are not negligible throughout the collapse process. Hence there is no guarantee that either approximation can reproduce the correct features of the collapse process (cf. Shandarin et al. 1995). In particular, we do not know from these results whether the NMA or LTA would produce pancake collapse given the initial conditions we have chosen. We also do not know which approximation will be more accurate for generic initial conditions, although Figure 1 suggests that it may be better to neglect M_{ij} than H_{ij}.

5.3 Pancakes Versus Spindles

The oblate and prolate configurations are distinguished by the signature of the eigenvalues of E_{ij} and σ_{ij}. For the collapsing oblate (pancake) configuration, the eigen-

values of E_{ij} have the signature $(-,-,+)$ and those of σ_{ij} have $(+,+,-)$. For the collapsing prolate (spindle) configuration, E_{ij} has eigenvalues with signature $(-,+,+)$ and σ_{ij} has $(+,-,-)$. One way to see why this is true is by inspecting equations (5.8) and (5.9). For the pancake configuration, one can use the fact that α_3 is close to 2 (supposing collapse occurs in the third direction) while α_1 and α_2 almost vanish. For the spindle configuration, suppose that collapse occurs for the second and third direction and suppose for simplicity that they collapse at the same rate. Then from equation (5.2), one can show that close to the spindle configuration, $\alpha_3 \simeq \alpha_2 \simeq 1$ and $\alpha_1 \simeq 0$. Using this and equations (5.8) and (5.9), it is possible to obtain the signature for the eigenvalues of E_{ij} and σ_{ij}. Note also that it is sufficient to consider only the divergent parts of E_{ij} and σ_{ij} to get the right signatures.

Consider equation (4.25). This is the tidal evolution equation of the NMA, which ignores the magnetic part of the Weyl tensor. First of all, the term proportional to a'/a always tends to decrease E_{ij}, encouraging spherical collapse. But by the time the motion of the object under consideration breaks away from the expansion of the universe, this term becomes unimportant. Suppose now that the object is close to the pancake configuration with E_{ij} having signature $(-,-,+)$ and σ_{ij} having $(+,+,-)$. Then it can be seen that all the terms favor pancake collapse (or favor neither pancakes nor spindles) except the shear-tide coupling terms $\delta_{ij}\sigma^{kl}E_{kl} - 3\sigma^{k}{}_{(i}E_{j)k}$. The net sign of these two terms is such that the growth of E_{ij} towards the pancake signature is suppressed. Suppose on the other hand that the object is close to the spindle configuration with E_{ij} and σ_{ij} having signatures $(-,+,+)$ and $(+,-,-)$ respectively. Then all the terms, including the shear-tide couplings, encourage the growth of tide towards the spindle signature. In other words, the NMA on the whole favors collapse toward the prolate or spindle configuration.

Consider, on the other hand, the exact tidal evolution equation (4.22). For an object with a very short third axis compared to the other two, we expect α_3 to be slightly less than but close to 2 and α_1 and α_2 to be small and positive. Substituting this into equation (5.10) and looking only at the most divergent terms, one can verify that M_{ij} has signature $(-,-,+)$ close to the pancake configuration. Using similar

arguments, it can be deduced that M_{ij} has signature $(-,+,+)$ close to the spindle configuration. Hence, M_{ij} has the same signature as E_{ij} close to collapse, whether it be pancake or spindle; therefore it stabilizes collapse just like the Hubble damping term proportional to a'/a. Hence ignoring M_{ij}, which is the LTA, does not favor spindles over pancakes. This is very different from the NMA.

Equation (4.23) tells us that M_{ij} contains both $-\nabla_k \epsilon^{kl}{}_{(i}H_{j)l}$ and the shear-tide coupling terms. We can now see what is wrong with the NMA — for the spindle configuration, $-\nabla_k \epsilon^{kl}{}_{(i}H_{j)l}$ has a signature that is opposite to the shear-tide coupling terms, and it is large enough to reverse the spindle-enhancing effect of the latter. As a result, M_{ij} as a whole, which includes the sum of these terms, plays no favorites.

Numerical integration bears out this analysis. We tested the LTA and NMA by integrating the sets of local Lagrangian fluid equations which are obtained by ignoring the relevant terms in the tidal evolution equation: equations (4.11), (4.12), (4.14), and either (4.25) or (4.26). The tensor equations were diagonalized along principal axes. Initial conditions were chosen using equations (5.4) and (5.7)–(5.9) so as to correspond to the homogeneous ellipsoid model with initial axis ratios $1 : 1.25 : 1.5$. Given the numerical solution for $\theta(\tau)$ and $\sigma_{ij}(\tau)$, we then predicted the evolution of the homogeneous ellipsoid axis lengths by solving equations (5.7)–(5.8) for R_i'/R_i and numerically integrating it to get $R_i(\tau)$. For comparison, we also computed the prediction of the ZA for the axis evolution given the same initial conditions. We obtained the same results for the ZA by integrating the local Lagrangian fluid equations (4.17) and (4.18) as we did from equation (5.4).

Figure 2 compares the local approximations (ZA, LTA, NMA) for the evolution of the axis lengths with each other and with the solution given by integrating equations (5.1). Both the ZA and LTA reproduce the qualitative features of pancake collapse. As we have already noted, the NMA predicts collapse to a spindle instead of pancake. For these initial conditions, at least, the LTA is even more accurate than the ZA. The LTA overestimates the expansion factor at collapse by only 3%, compared with 52% for the ZA. The LTA appears to rectify one of the well-known problems with the ZA, namely the fact that it underestimates the rapidity of collapse for non-planar

perturbations. This result is consistent with our observation in §4.3 that the LTA is exact for spherical and cylindrical symmetry. We note in passing that the modified Zel'dovich approximation mentioned in the introduction performs even a little better than the LTA for the above test case but it suffers from unphysical singularities for more generic initial conditions (Reisenegger & Miralda-Escudé 1995).

It is also useful to compare the local approximations with the exact solution for cylindrically symmetric perturbations. Consider a homogeneous overdense cylinder in an Einstein-de Sitter universe, with radius $R(t)$. The equation of motion is given by Fillmore & Goldreich (1984). It can be written in a form corresponding to equation (5.1):

$$\frac{d^2R}{dt^2} = -2\pi GR\left[\frac{2}{3}\rho_b + (\rho_c - \rho_b)\right] , \tag{5.11}$$

where ρ_c is the density inside the cylinder. In fact, this is identical to equation (5.1) for a homogeneous ellipsoid with axis ratios $R : R : \infty$, for which $\alpha_1 = \alpha_2 = 1$, $\alpha_3 = 0$. We repeated the comparison of local approximations with the exact solution given by integrating equation (5.11). The results are shown in Figure 3. The LTA is exact, while the NMA underestimates the expansion factor at collapse (by 23%) and the ZA overestimates it (by 36%).

An exact solution also exists for a homogeneous ellipsoid in a vacuum (non-expanding) background (Lin et al. 1965). It is easy to modify the NMA and LTA equations for this case, by setting $a = 1$ and $\rho_b = \bar{\rho} = 0$. We did not integrate the non-cosmological analog of the ZA. As for Figure 2, we set the initial axis ratios to be $1 : 1.25 : 1.5$, although in this case we set the initial velocity field to zero. Figure 4 shows the results. Once again we see that the LTA is rather accurate for generic initial conditions (the collapse time here is 1.5% too large) and leads to pancake collapse, while the NMA incorrectly predicts spindle collapse.

To compare the three local approximations (ZA, NMA, LTA) with more general initial conditions, we follow the notations of Bertschinger & Jain (1994) and write

traceless symmetric tensors in terms of a magnitude and an angle:

$$\sigma_{ij} = \frac{2}{3}\,\sigma\,Q_{ij}(\alpha)\ , \quad E_{ij} = \frac{8\pi}{3}\,G\bar{\rho}a^2\,\epsilon\,(1+\delta)\,Q_{ij}(\beta)\ . \tag{5.12}$$

We have introduced new scalars $\sigma \leq 0$, $\epsilon \geq 0$, α and β ($0 \leq \alpha, \beta \leq \pi$), and a one-parameter traceless quadrupole matrix

$$Q_{ij}(\alpha) \equiv \mathrm{diag}\left[\cos\left(\frac{\alpha+2\pi}{3}\right), \cos\left(\frac{\alpha-2\pi}{3}\right), \cos\left(\frac{\alpha}{3}\right)\right]\ . \tag{5.13}$$

With this parametrization, oblate configurations have $\cos\alpha > 0$ while prolate configurations have $\cos\alpha < 0$. Of course, the shape of a perturbation can change with time. The equations of motion for σ, ϵ, α, and β for the NMA are given by Bertschinger & Jain. For the LTA, their equations (13) and (14) are changed to become

$$\frac{d\epsilon}{d\tau} - \theta\epsilon = -\sigma\cos\left(\frac{\alpha-\beta}{3}\right)\ , \tag{5.14}$$

$$\frac{d\beta}{d\tau} = -\frac{3\sigma}{\epsilon}\sin\left(\frac{\alpha-\beta}{3}\right)\ . \tag{5.15}$$

One quantity of interest for general initial conditions is the expansion factor at collapse, i.e., the linear overdensity when a given mass element collapses. Following Bertschinger & Jain (1994), we parametrize the initial conditions by ϵ_0 and α_0, which are related to the values of ϵ and α in linear theory through $\epsilon = a\epsilon_0$ and $\alpha = \alpha_0$. Because initially underdense perturbations can collapse if the shear is sufficiently strong, we treat both initially overdense and underdense perturbations by specifying $\delta_0 = \pm 1$, respectively (δ_0 being related to δ in linear theory by $\delta = a\delta_0$). The expansion factor at collapse, a_c, is determined by integrating the local evolution equations for the LTA and NMA. For the ZA, it is simpler to use equations (4.4) and (4.8), noting that collapse occurs when the determinant of $\partial x^i/\partial q_j$ vanishes. With our parametrization of the initial velocity and gravity gradient tensors, it follows that

$$a_c = \frac{3}{\delta_0 + 2\epsilon_0\cos(\alpha_0/3)} \quad \text{(ZA)}\ . \tag{5.16}$$

The collapse expansion factor a_c is defined to be the absolute value of the linear overdensity when a given mass element collapses to infinite density. For example, an overdense spherical perturbation collapses when $a_c = 1.686$, while a cylindrical perturbation collapses when $a_c = 1.466$ and a plane-parallel perturbation collapses when $a_c = 1$. Although there exists no exact solution for arbitrary initial conditions, it is informative to compare all three methods. Based on our previous results we expect the LTA to be accurate to a few percent.

We plot contours of constant collapse time for different initial tidal parameters ϵ_0 and α_0 for the three local approximations in Figure 5. In each part, the left panel gives results for overdense perturbations while the right panel is for initially underdense perturbations. Figure 5b presents the same results as Figures 1 and 2 of Bertschinger & Jain (1994). We see that the LTA and ZA are qualitatively similar, although the ZA overestimates the collapse time for overdense configurations with small tide (near the center of the figures). According to the ZA, $a_c = 3$ for spherical perturbations while the exact value is $\frac{5}{3}(2/3\pi)^{2/3} = 1.68647\ldots$. Both the ZA and LTA indicate more rapid collapse for initially oblate configurations. As noted by Bertschinger & Jain, initially prolate configurations collapse faster in the NMA because according to its incorrect dynamics initially oblate configurations must change shape before collapsing to a spindle.

Bertschinger & Jain (1994) also noted that shear can lead to collapse of underdense perturbations. From Figure 5, we see that that the size of the non-collapsing region in parameter space (in the middle of the right-hand panels) is largest for the NMA and smallest for the ZA, indicating that the NMA underestimates the fraction of initial underdense perturbations that can collapse, while the ZA overestimates it. Using the probability distribution of ϵ_0 and α_0 derived by Bertschinger & Jain for a Gaussian random field, we find that the probability that a randomly chosen mass element will collapse is 0.780 for NMA, 0.888 for LTA, and 0.920 for ZA. Thus, taking the LTA as the most accurate, approximately 78% ($= 2 \times 0.888 - 1$) of the initially underdense perturbations (and 100% of the overdense ones) will collapse. This estimate neglects the crossing of mass elements, which increases the likelihood of collapse

by increasing the density owing to multi-streaming Indeed, we expect most mass elements to collapse eventually in a perturbed self-gravitating cold dust medium. Note also that the probabilities quoted above apply to underdense regions in the early universe, not the universe today. For instance, the fact that most underdense mass elements in the early universe collapse does not imply the probability of finding voids in the present universe is small. On the contrary, weighing by volume, the probability of locating voids or highly underdense regions in a random sample of the universe today is quite high because most of the mass has collapsed, leaving most of the volume emtpy.

5.4 Conclusion

In Chapters 4 and 5, we have discussed three different local approximations for gravitational collapse of perturbations in an expanding universe: the Zel'dovich approximation (ZA), the non-magnetic approximation (NMA) of Bertschinger & Jain (1994), and a new local tidal approximation (LTA) introduced here. Conventionally, the ZA is presented as a mapping of Lagrangian to Eulerian positions. However, we showed that it can also be regarded as a certain truncation of the set of Lagrangian fluid equations for the density, velocity gradient, and tide following a fluid element of cold dust. With the ZA, the gravity gradient is explicitly proportional to the velocity gradient, resulting in modifications to the Raychaudhuri and shear evolution equations. The tidal evolution equation need not be integrated in the ZA because the gravity field acting on a mass element is given by a simple extrapolation of initial conditions. The other two approximations we discuss extend the ZA by integrating the exact Raychaudhuri and shear evolution equations, with approximations made only to the tidal evolution equation.

All three local approximations are exact for plane-parallel perturbations. However, the behavior for other shapes of perturbations shows significant differences in behavior. The ZA is only approximate for non-plane-parallel distributions. The NMA is exact for spherical perturbations but not cylindrical ones. The LTA is exact for

spherical and cylindrical perturbations and, more generally, for any growing-mode perturbations whose gravitational equipotential surfaces have constant shape with time.

In order to test these approximations for non-symmetrical shapes, we compared them in the case of the collapse of a homogeneous ellipsoid. As expected from the results of Bertschinger & Jain (1994), we find that the NMA generically produces spindle-like singularities at collapse. The LTA, on the other hand, generically produces pancakes, just like the ZA. For triaxial ellipsoids, we compared numerical integrations of the local evolution equations with known solutions for a homogeneous ellipsoid in both cosmological and vacuum backgrounds. (An exact solution exists for the latter case while, in the former case, the homogeneous ellipsoid solution is not really exact because tides will cause the background, and then the ellipsoid itself, to become inhomogeneous. However, these effects are expected to be small.) We find that the LTA is significantly more accurate than the ZA (see Fig. 2).

These results suggest we have found a promising new approximation for nonlinear gravitational instability. However, we have only studied the evolution of isolated irrotational perturbations. Caution is needed because we do not know how accurate the LTA is for more general initial conditions, for example, those with vorticity. Moreover, we do not know by how much the tide produced by other mass elements degrades the accuracy. External tides modulate the equipotentials surrounding a mass element; qualitatively, we expect little effect as long as the external tide evolves weakly or is small compared with the trace part of the gravity gradient. Quantitative analysis is best done using N-body simulations, which we leave for later work.

The LTA has one significant limitation compared with the ZA. It tells us only the internal state of a given mass element (density, expansion rate, shear) and the tide on the element, but does not give the position of the element. However, for many purposes one cares more about the internal evolution of a mass element than about its position. For example, simple models of galaxy formation are based on spherical infall. These can be improved by inclusion of shear and tides (Bond & Myers 1993; Eisenstein & Loeb 1995). Our approximations could lead to even

more accurate models of this sort. Also, if one does need to know the positions of mass elements, then one can always supplement the Lagrangian fluid equations by the equation of motion for positions, perhaps using the Zel'dovich approximation or higher-order Lagrangian equations of motion. In principle, by following the velocity gradient for many mass elements, one can reconstruct the velocity field (up to an irrelevant overall constant), and then integrate the positions with $d\vec{x}/d\tau = \vec{v}$. An equivalent procedure was suggested by Matarrese et al. (1993).

Perhaps the most important reason for seeking new approximations like the LTA is that we still lack a good understanding of the behavior of collisionless systems under nonlinear gravitational instability. Future work will tell whether local Lagrangian flow methods will provide new insights.

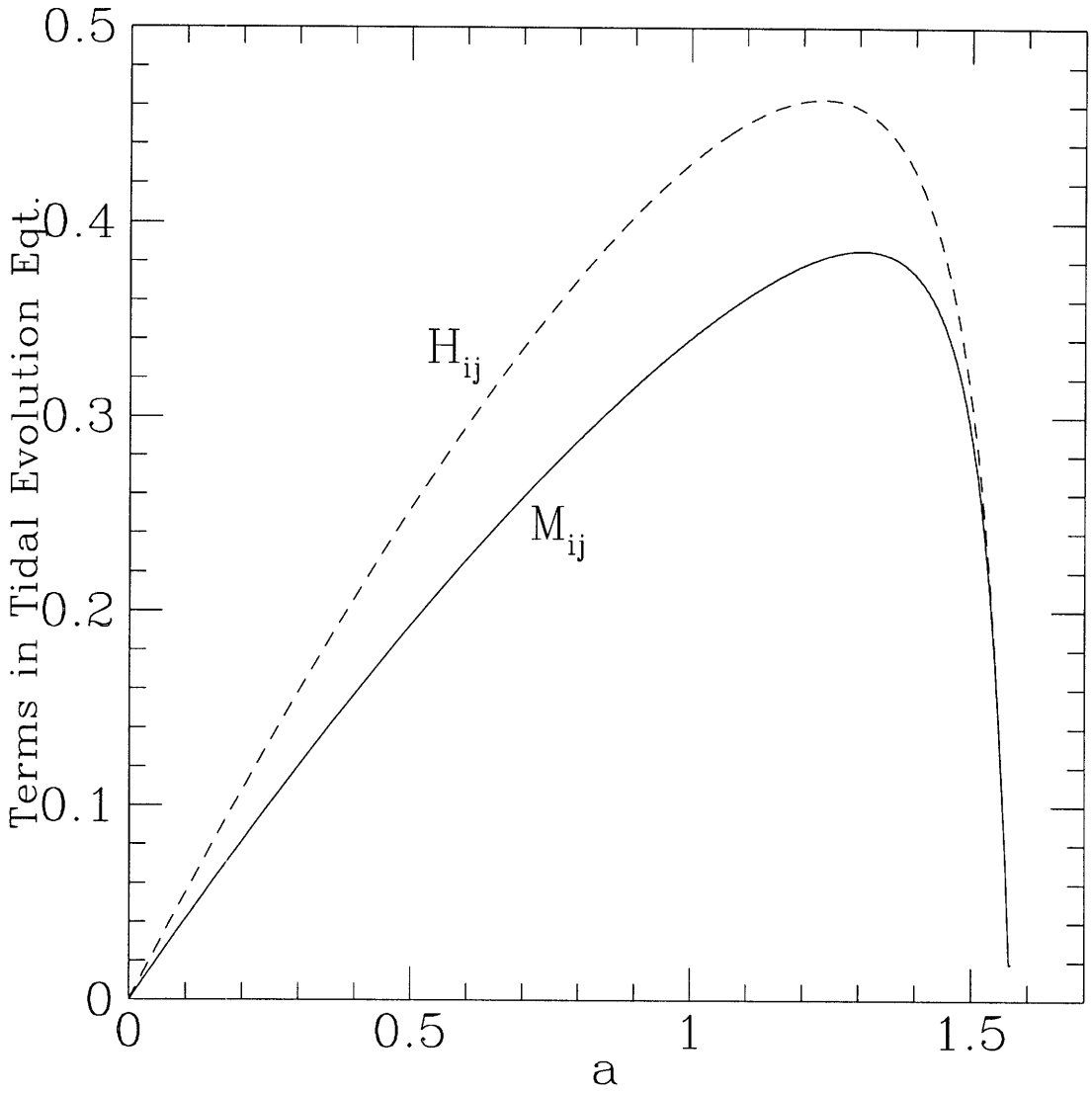

Fig. 5-1.— Evolution of the magnitudes of $-\nabla_k \epsilon^{kl}{}_{(i} H_{j)l}$ (dashed line) and M_{ij} (solid line) divided by the magnitude of $4\pi G a^2 \rho \sigma_{ij}$, evaluated for a homogeneous ellipsoid with initial axis ratios $1 : 1.25 : 1.5$ embedded in an expanding universe. The magnitude of a matrix is defined here as the square root of sum of squares of eigenvalues.

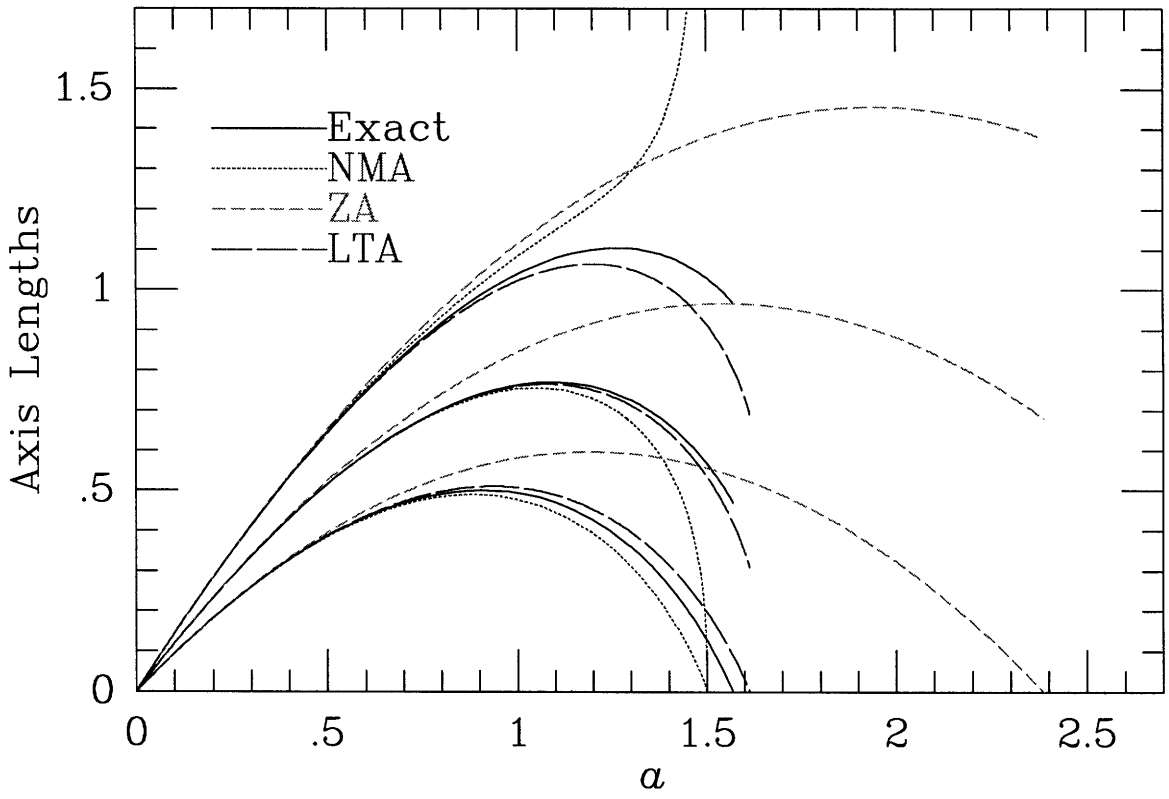

Fig. 5-2.— The evolution of axis lengths for a homogeneous ellipsoid embedded in an expanding universe. The initial axis ratios are $1 : 1.25 : 1.5$. The "exact" solution (ignoring development of inhomogeneity, solid curve) is compared with the ZA (short dashed curve) and two local approximations: the local tidal approximation (LTA, long dashed curve) and the non-magnetic approximation of Bertschinger & Jain (NMA, dotted curve).

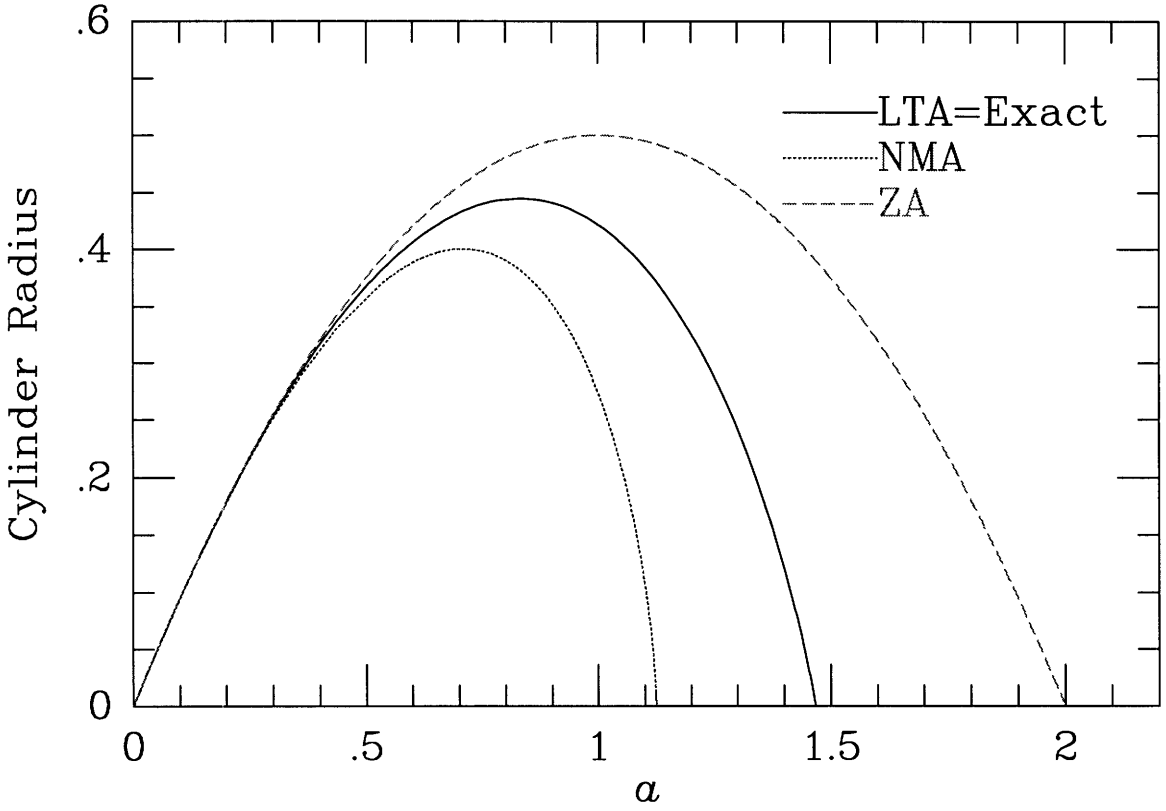

Fig. 5-3.— The evolution of the radius of a cylindrical perturbation in an expanding universe, corresponding to a homogeneous ellipsoid with axes $R : R : \infty$ (a cylinder). The exact solution (solid curve) is compared with the ZA (short dashed curve) and the NMA (dotted curve). The LTA is exact for this case.

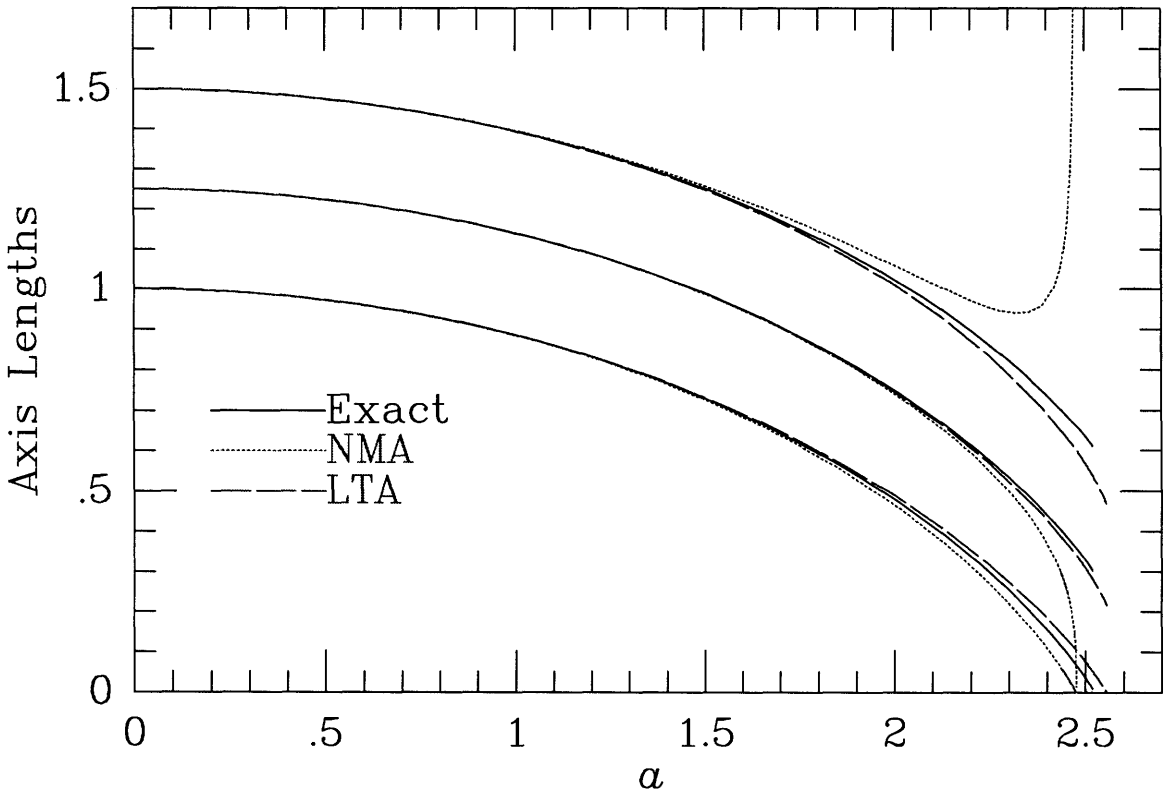

Fig. 5-4.— The evolution of axis lengths for a homogeneous ellipsoid embedded in empty space. The initial axis ratios are 1 : 1.25 : 1.5. The exact solution (solid curve) is compared with the predictions of LTA (long dashed curve) and NMA (dotted curve).

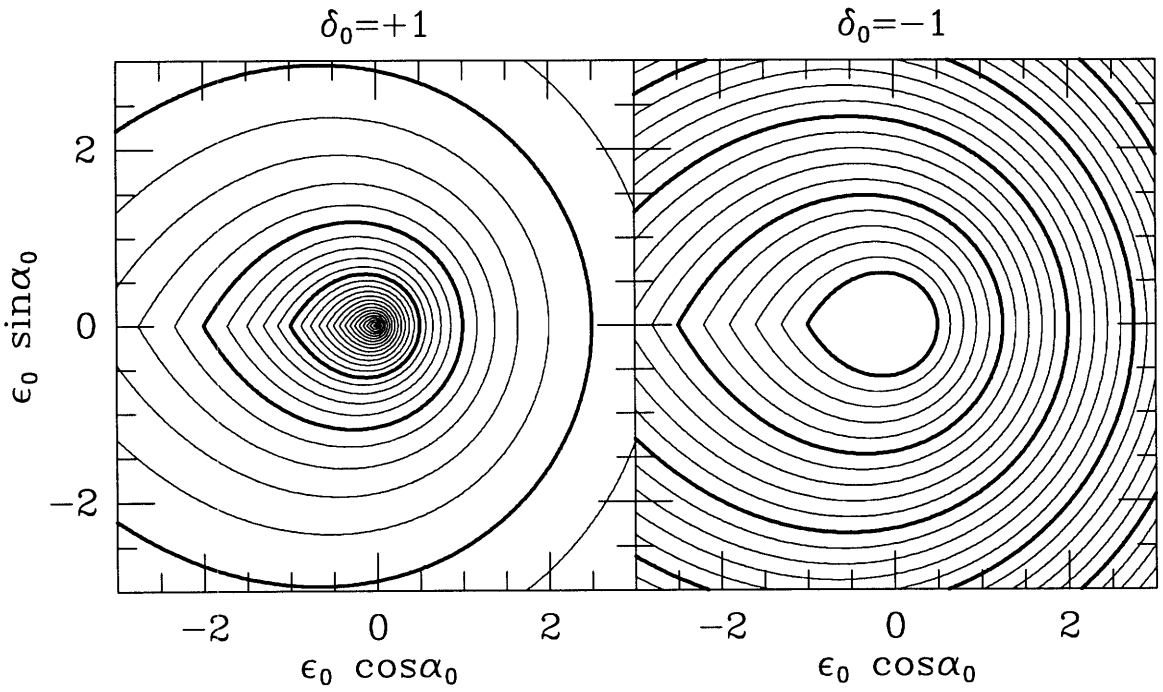

Fig. 5-5a.— Contours of constant collapse time computed using the ZA, expressed by the cosmic expansion factor a_c or its reciprocal, versus initial tidal field parameters. Left panel: initial positive density perturbations. The light (heavy) contours are spaced by 0.1 (0.5) in a_c, with the outermost contour $a_c = 0.4$ and the central value (corresponding to spherical collapse) $a_c = 3.0$. The ZA significantly overestimates the collapse time for low-shear perturbations. Right panel: initial negative density perturbations. The light (heavy) contours are spaced by 0.1 (0.5) in a_c^{-1}, with the innermost contour $a_c^{-1} = 0$ and the outermost one $a_c^{-1} = 2.3$. Initial perturbations in the central region do not collapse. Perturbations are oblate (prolate) for $\epsilon_0 \cos\alpha_0 > 0$ ($\epsilon_0 \cos\alpha_0 < 0$).

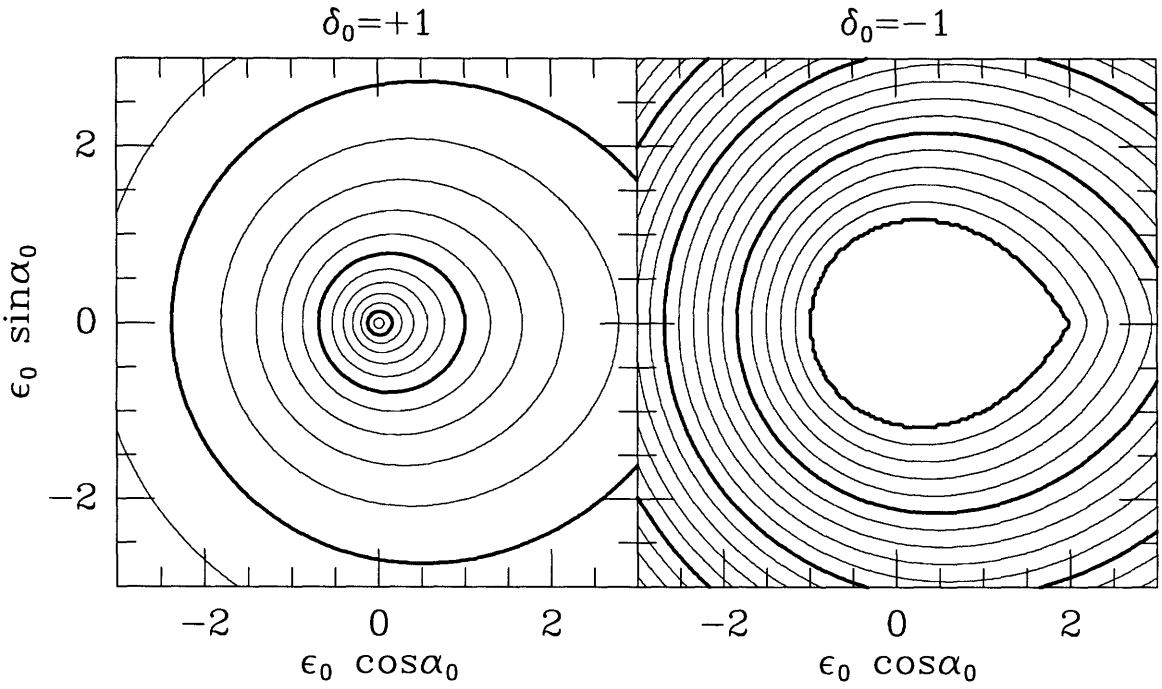

Fig. 5-5b.— Same as Fig. 5a except that the NMA is used. In the left panel the innermost contour is $a_c = 1.6$. In the right panel the outermost contour is $a_c^{-1} = 1.8$. The smaller extent of the contours for prolate configurations ($\epsilon_0 \cos\alpha_0 < 0$) reflects the fact that the NMA favors prolate collapse.

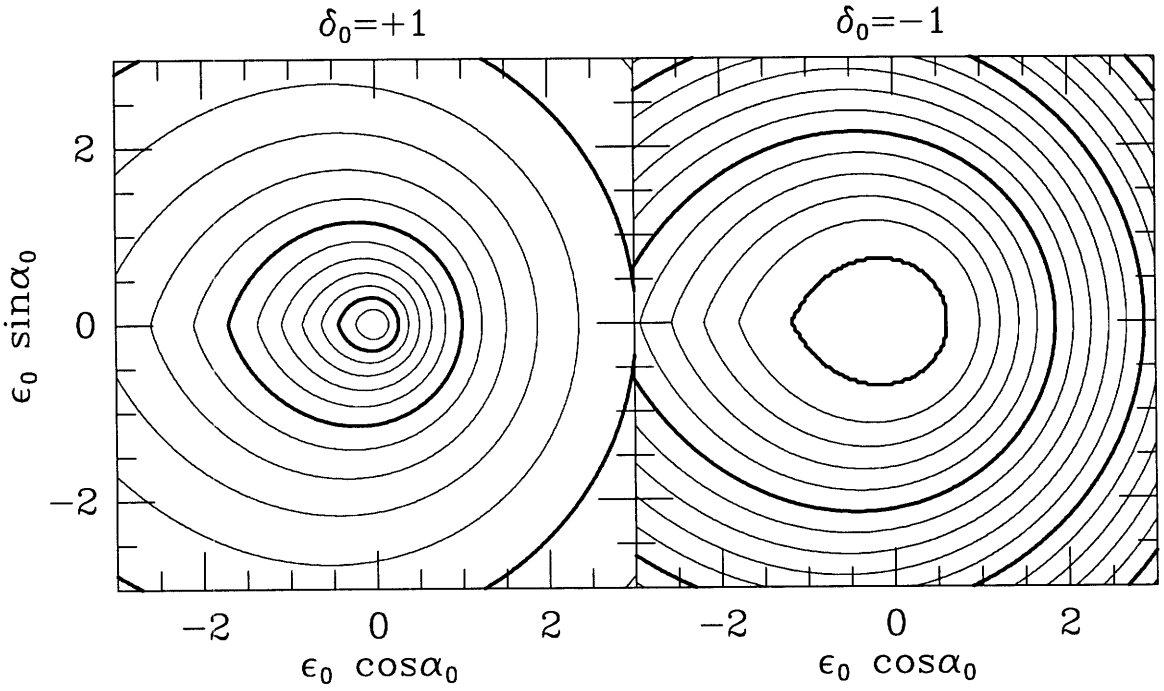

Fig. 5-5c.— Same as Fig. 5a except that the LTA is used. In the left panel the innermost contour is $a_c = 1.6$. In the right panel the outermost contour is $a_c^{-1} = 1.6$. The LTA, like the ZA, favors oblate (pancake) collapse over prolate (spindle) collapse.

Chapter 6

The Statistics of Density Peaks and the Column Density Distribution of the Lyman-Alpha Forest

6.1 Introduction

There is a long history of theoretical efforts to place the study of the Lyman-α forest within the framework of cosmological structure formation theories (Rees 1986; Bond et al. 1988; McGill 1990; Bi et al. 1992). Recent work making use of numerical simulations has greatly advanced our understanding in this direction (Cen et al. 1994; Zhang et al. 1995; Hernquist et al. 1995; Petitjean et al. 1995; Miralda-Escudé et al. 1995). (See also Bi et al. 1995 for a linear theory calculation). The emerging picture is that it is possible to account for all the observed properties of the Lyman-α forest (with column densities [$N_{\rm HI}$] less than about $10^{17}\,{\rm cm}^{-2}$) by assuming it originates from the small scale structure, including the network of filaments, pancakes and mild density fluctuations, which arises naturally in hierarchical clustering cosmological models (Weinberg et al. 1996).

A commonly used statistic to characterize the forest is its column density distribution, the number of absorption lines per unit column density per unit redshift as a function of column density. Other useful statistics include line-line correlations and

the distributions of b-values and equivalent widths (Murdoch et al. 1986; Carswell et al. 1991; Press et al. 1993; Cristiani et al. 1995). There have also been proposals of new statistical tools (Meiksin & Bouchet 1995; Miralda-Escudé et al. 1995; Pando & Fang 1996). (See Tytler 1992 for a general overview of the statistical issues concerning quasar absorption systems.) We focus our attention on the column density distribution in the present work.

One of the most striking features of the observed column density distribution of quasar absorption systems is that it can be approximated by a single power law that extends over many orders of magnitude. This was emphasized by Tytler (1987), among others who found that in the range $10^{13} < N_{\rm HI} < 10^{22}\,{\rm cm}^{-2}$, the distribution was reasonably well represented by a power law, $\propto N_{\rm HI}^{-\beta}$ with $\beta = 1.51 \pm 0.02$. However, there exists evidence of at least one break. It has been demonstrated that there is a deficit of absorption systems somewhere in the column density range 10^{14} to $10^{17}\,{\rm cm}^{-2}$ compared to a power-law extrapolation of the distribution from lower column densities (Carswell et al. 1987; Petitjean et al. 1993; Hu et al. 1995; Giallongo et al. 1996). For reasons that have to do with the nature of the approximations that we make (Sec. 6.5.1), we focus our attention on absorption systems with column densities in the range $10^{12.5} < N_{\rm HI} < 10^{14.5}\,{\rm cm}^{-2}$. Hu et al. (1995) obtained $\beta = 1.46$ with a 95% confidence range of $(1.37, 1.51)$ in the column density range $10^{12.3} < N_{\rm HI} < 10^{14.5}\,{\rm cm}^{-2}$. Lu et al. (1996) found the same best-fit β of 1.46 for the same range of column densities.

An obvious ultimate goal of recent theoretical work on the Lyman-α forest is to constrain theories of structure formation. The natural question is: what determines the normalization and slope of the column density distribution? What are the major determining factors, in addition to the usual parameters specified by a given cosmological model? To answer these questions, another question has to be addressed: what are the analytical and/or computational tools necessary to make accurate predictions for the column density distribution, given all the required parameters?

Accordingly, the present work can be divided into three parts, where the tools are developed, the factors that influence the column density distribution are analyzed

and one application to a class of cosmological models is discussed.

Numerical hydrodynamic simulations (Cen et al. 1994; Zhang et al. 1995; Hernquist et al. 1995; Miralda-Escudé et al. 1995) provide the most obvious tools to study the Lyman-α forest. Computational costs, however, prevent one from testing a lot of cosmological models. The Zel'dovich approximation (Zel'dovich 1970), with appropriate smoothing, is presented in this chapter as an efficient and accurate alternative. Our basic assumption is that the part of the Lyman-α forest with column densities less than about $10^{14.5}\,\mathrm{cm}^{-1}$ arises mostly from regions of mild overdensities ($\lesssim 5$) or even underdensities which have not undergone orbit-crossing. The Zel'dovich approximation can be coupled with the equations governing the thermal and ionization states of the gas to yield predictions for the density of neutral hydrogen and the peculiar velocity as a function of position. Absorption spectra are then generated and analyzed. Basic expressions for the absorption optical depth are presented in Sec. 6.2 and the approximations that go into its computation are discussed in Sec. 6.3.

Given an absorption spectrum, the column density distribution depends on the methods of identifying lines and assigning column densities. This is discussed in Sec. 6.4.1. Using a method described in that section (Miralda-Escudé et al. 1995), we investigate the effects of peculiar velocities on the column density distribution. We find that although peculiar velocities can strongly influence the shapes of absorption profiles, they play a minor role in determining the column density distribution. The various interesting effects of peculiar velocities are discussed in Sec. 6.4.2. Motivated by this finding, a very different way of assigning column densities is introduced in Sec. 6.5, in which no absorption spectrum needs to be generated. In the absence of peculiar velocities, there is a one-to-one correspondence between density peaks in real space (if they are separated by a distance larger than a minimum corresponding to the thermal broadening width) and minima in the absorption spectrum. One can simply identify density peaks in real space and assign a column density to each based on its height and curvature. The number density of such peaks is then converted into the column density distribution. We apply this procedure (we call it the Density-Peak-Ansätz) to the density field predicted by the truncated Zel'dovich approximation and

test the result against that of a full hydrodynamic simulation. The column density distribution obtained in this way is compared to that obtained from the hydrodynamic simulation using the Voigt-profile-fitting-technique, which is the line-identification method most commonly used. The level of agreement is found to be excellent. In Sec. 6.5.1, we discuss the range of parameters in which our computed column density distribution is expected to be reliable.

Armed with the right tools, we turn our attention to the second question: what factors determine the column density distribution? They can be divided into two categories. One has to do with properties of the intergalactic medium, including its temperature, the equation of state, the ionizing radiation intensity and the baryon density. Uncertainties in all of them have to be taken into account before one can meaningfully confront theories with observations. We distinguish between the factors that mostly affect the normalization of the column density distribution and those that mostly affect its slope. It is also emphasized that the temperature and the equation of state depend on the reionization history of the universe (see Hui and Gnedin 1996). The second set of factors affecting the column density distribution has to do with the specific cosmological model, namely the normalization and shape of the corresponding power spectrum. We study a few Cold Dark Matter (CDM) models in Sec.6.6 for this purpose. It is found that the amount of (linear) power on comoving scales of around $2\,h_{100}$ Mpc^{-1} to $20\,h_{100}$ Mpc^{-1} is the single most important factor in determining the slope of the column density distribution. Increasing the power tends to flattens the distribution. In Sec.6.7, we write down expressions relating the slope of the column density distribution to the equation of state and the amount of power on small scales.

We then study a group of Cold plus Hot Dark Matter (CHDM) models in Sec. 6.8, making use of the insights gained in Sec. 6.6. The $\Omega_\nu = 0.2$ (density parameter in neutrino) models have steeper column density distributions compared to those with $\Omega_\nu = 0.1$ because they have less power on the relevant scales. In particular, the low Hubble constant ($H_0 = 50\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$) $\Omega_\nu = 0.2$ models predict slopes that are steeper than the observed one for most of the parameter-space specifying the properties of the intergalactic medium. Only for almost isothermal equations of state

can the two be made consistent with each other. We emphasize however that a more detailed comparison between the models and observations, taking fully into account instrumental noise and biases of the line-identification method(s), is necessary before one can firmly reject a model. In Sec. 6.9, we summarize our conclusions.

On our notations, bold faced letters are reserved for three-dimensional vectors. The symbols $\mathbf{v}_{\rm pec}$ and $\mathbf{x}$ denote the three-dimensional peculiar velocity and comoving position while $v_{\rm pec}$ and x are their counterparts along the line of sight of interest. Standard symbols are used for cosmological parameters: H for the Hubble constant as a function of z, H_0 for the Hubble constant today, h_{100} for $H_0/100$ kms^{-1}Mpc^{-1}, Ω_0 for the density parameter today, with the subscript b to denote its baryon portion and ν its neutrino content. The term multiple-streaming is reserved for the situation where a single observed redshift corresponds to more than one position in real space. We distinguish it from the term orbit-crossing, which is commonly used interchangeably with multiple-streaming in other contexts. Orbit-crossing refers to the case where a single position has more than one velocity.

This chapter is largely based on a preprint (Hui, Gnedin and Zhang 1996). We thank Edmund Bertschinger for generously providing computer codes which, after slight modifications, are used to generate the Zel'dovich density and velocity fields.

6.2 Cosmological Lyman-Alpha Absorption in a Fluctuating Medium: Basic Results

A photon emitted with energy higher than 10.196 eV (wavelength of 1216 $\AA$) by a distant quasar is continuously redshifted as it travels through the intergalactic medium until it reaches the observer. At some intermediate point, the photon is redshifted to around 1216 $\AA$ in the rest frame of the intervening medium, which may contain neutral hydrogen. It excites the Lyman-α transition and is absorbed. Consider a particular line of sight from the observer to the quasar, the optical depth (related to the probability of transmission by $e^{-\tau}$) of a photon at a given observed frequency ν_O

is given by:

$$\tau(\nu_O) = \int_{x_A}^{x_B} n_{\rm HI}\sigma_\alpha \frac{dx}{1+z}, \tag{6.1}$$

where x is the comoving radial coordinate of some intermediate point along the line of sight, z is the redshift and $n_{\rm HI}$ is the proper number density of neutral hydrogen at that point. The limits of integration, x_A and x_B, are the comoving positions of the observer and the quasar. The Lyman-α absorption cross-section is denoted by σ_α. It is a function of the frequency of the photon with respect to the rest frame of the intervening hydrogen at position x. Let us call this frequency ν. The cross-section is peaked when ν is equal to the Lyman-α frequency, ν_α.

The frequency ν is related to the observed frequency ν_0 by:

$$\nu = \nu_0(1+z)\left(1+\frac{v_{\rm pec}}{c}\right), \tag{6.2}$$

where $v_{\rm pec}$ is the peculiar velocity at position x and $1+z$ is the redshift factor *due to the uniform Hubble expansion alone* at the same position. The peculiar velocity of the observer, which merely displaces the whole spectrum by a constant amount (independent of x), is ignored. The quantity v_{pec}/c, where c is the speed of light, is much smaller than 1.

It proves convenient for later discussion to expand z around some mean redshift of interest $\bar{z}$, which could be the redshift of a simulation output or the average redshift of an observed spectrum with limited redshift range. Using $dx = cdt/a$, where a is the Hubble scale factor and t is the proper time, it can be shown that

$$\nu = \nu_0(1+\bar{z})\left(1+\frac{u}{c}\right), \; u \equiv \frac{\bar{H}}{1+\bar{z}}(x-\bar{x}) + v_{\rm pec}(x), \tag{6.3}$$

where $\bar{x}$ is the position at which the redshift due to Hubble expansion coincides exactly with $\bar{z}$. The Hubble constant at the same redshift is denoted by $\bar{H}$. We assume the range of x is small enough so that $u/c \ll 1$. The convention $a = 1$ today is adopted.

The velocity coordinate u defined above contains contributions from both the Hubble expansion and the peculiar motion. Without peculiar motion, u increases

monotonically with x and is in fact linear in x. Peculiar velocities destroy the linear relation and could give rise to situations where a given u corresponds to more than one position x. It implies that a photon of a given observed frequency ν_0 can have the same rest-frame frequency ν at more than one place in its trajectory from the quasar to the observer. We reserve the term multiple-streaming to this situation and distinguish it from orbit-crossing where a given x carries more than one $v_{\rm pec}$ or u. We will return to the subject of multiple-streaming in Sec. 6.4.2 and that of orbit-crossing in Sec. 6.3.1 .

We define one more velocity coordinate u_O, which is related to the observed frequency ν_O by:

$$\nu_O = \frac{\nu_\alpha}{1+\bar{z}}\left(1 - \frac{u_O}{c}\right) \tag{6.4}$$

where ν_α is the Lyman-α frequency. The velocity coordinate u_0 is simply equal to u when ν coincides exactly with ν_α (this can be seen by comparing equations [6.3] and [6.4], bearing in mind that u/c and u_0/c are both assumed to be much less than 1).

With the definitions in place, we change the variable from x to u in equation (6.1), which results in the following expression for τ, now a function of u_O:

$$\tau(u_O) = \sum \int_{u_A}^{u_B} \frac{n_{\rm HI}}{1+z}\left|\frac{du}{dx}\right|^{-1} \sigma_\alpha du \ , \ \sigma_\alpha = \sigma_{\alpha 0}\frac{c}{b\sqrt{\pi}} e^{-(u-u_O)^2/b^2} \ . \tag{6.5}$$

The summation refers to a sum over multiple-streams (all the x's within the range $x_A - x_B$ that corresponds to a given u), and $n_{\rm HI}$, z and $|du/dx|^{-1}$ are now functions of u. The limits of integration u_A and u_B are the velocity coordinates corresponding to the positions x_A and x_B (assuming no orbit-crossing so that each x carries one u). Note that in practice, only a limited range of u contributes to τ for a limited range of u_O so that one can replace the redshift z with $\bar{z}$. The same is also true for equation (6.1).

The Lyman-α cross-section σ_α is expressed as a function of $u - u_O$. The constant $\sigma_{\alpha 0}$ is equal to the combination of fundamental physical constants $0.416\pi e^2/(m_e c \nu_\alpha)$, where e is the charge of an electron and m_e is its mass. It is about $4.5 \times 10^{-18}\,{\rm cm}^2$. The parameter b is equal to $\sqrt{2k_B T/m_p}$ where k_B is the Boltzmann constant, m_p is

the mass of a proton and T is the temperature of the gas.

The form of the line profile function above takes into account thermal broadening but ignores the natural line width. A more general profile function involves a convolution of the two, resulting in the Voigt profile (Rybicki & Lightman 1979). However, for column density less than about $10^{17}\,\mathrm{cm}^{-2}$, the simple thermal profile is adequate. Only for systems where the optical depth greatly exceeds one is it important to use the full Voigt profile. The reader is referred to Spitzer (1978) and Press and Rybicki (1993) for discussions of curve of growth analysis.

Note also that it is sometimes assumed b contains a component due to turbulent motion. We do not include it explicitly in our formalism. Bulk motion, on the other hand, is accounted for by v_{pec} or u.

Let us consider two different limits of equation (6.5).

Suppose there is a high local maximum in $n_{\mathrm{HI}}|du/dx|^{-1}$ at some $u = u_{max}$ with width in velocity space much smaller than the thermal width b. Then one can take the line profile function associated with σ_α out of the integral in equation (6.5) because $n_{\mathrm{HI}}|du/dx|^{-1}$ varies much more rapidly than the thermal profile:

$$\tau(u_O) = \left(\int_{\mathrm{max}} n_{\mathrm{HI}}(x)\frac{dx}{1+z}\right)\sigma_{\alpha 0}\frac{c}{b\sqrt{\pi}}e^{-(u_O-u_{\mathrm{max}})^2/b^2}\,, \tag{6.6}$$

where the variable of integration has been changed back from u to x. The equation holds when u_O is close enough to u_{max}. The integral is over the local maximum, assuming that the amount of neutral hydrogen away from the maximum does not cause significant absorption (until another maximum is encountered). One then sees an absorption line with an exponential profile in optical depth. While the width of the line tells us about b, which is proportional to the square root of the temperature, the depth of the line provides information about both b and the column density, which is the integral inside the first pair of brackets on the right hand side. Let us call this the narrow-maximum-limit.

Consider another limit of the integral (equation [6.5]) in which $n_{\mathrm{HI}}|du/dx|^{-1}$ varies slowly with u. Suppose the scale of variation is much larger than the thermal

width. In this case, one can leave the line profile function inside the integral but take the rest outside:[1]

$$\tau(u_O) = \sum \frac{n_{\rm HI}}{1+\bar{z}} \left|\frac{du}{dx}\right|^{-1} c\sigma_{\alpha 0}\,. \tag{6.7}$$

The velocity dependent terms on the right hand side are evaluated at u_O. The profile function has been integrated out.

In the above limit, τ does not necessarily have the thermal profile around its maxima. We will call this the broad-maximum-limit. An extreme example is that of a homogeneous medium, which gives rise to featureless and uniform absorption (Gunn & Peterson 1965).

Conventional analysis of quasar spectra involves identifying those parts of the spectra that are due to the Lyman-α absorption and fitting them with superpositions of the Voigt profiles (of which the thermal profiles are a subset) until the residual signal is consistent with noise. This technique was motivated by the picture of the intergalactic medium as consisting of a smooth component which causes relatively little absorption and a set of clouds that satisfy the narrow-maximum-limit. For each cloud, the best-fit Voigt profile then yields its temperature and column density according to equation (6.6).

However, it is clear that not all maxima in τ necessarily satisfy the conditions leading to equation (6.6). In fact, according to most structure formation theories, there invariably exists fluctuations in the intergalactic medium on scales larger than the thermal width. In the broad-maximum-limit, the shape of a local maximum in optical depth is determined by the distribution of $n_{\rm HI}$ and $|du/dx|$ around it. Each maximum in τ can still be identified as an absorption line and one can even apply standard techniques and try to fit its shape with superposition of the Voigt profiles. Given the best-fit Voigt profiles, one can assign a b-value (width of the profile) and a column density to each but it is no longer true, for instance, that the b-value thus obtained is equal to $\sqrt{2k_BT/m_p}$. A reasonable question to ask is whether there are other practical methods of identifying absorption lines and assigning column

[1]The expression is not valid at velocity caustics, where du/dx vanishes. Further discussion can be found in Sec. 6.4.2.

densities without assuming every absorption line consists of a superposition of the Voigt profiles. This will be discussed in Sec. 6.4.1. One might even seek ways other than the column density distribution to describe the fluctuations in τ. See Miralda-Escudé et al. (1995) for a discussion of some of these methods.

In general, there are regions of high density and limited extent, galaxies for instance, which give rise to absorption profiles well approximated by the narrow-maximum-limit, but there are also regions in the intergalactic medium with gentle fluctuations where the broad-maximum-limit holds. Then there are those places where neither limit applies, in which cases, a full integration of equation (6.5) has to be carried out to compute the optical depth. To do so, one needs to know how the neutral hydrogen density, peculiar velocity and temperature vary with space. This is the subject of the next section. In any case, the above discussion should make it clear that the quasar absorption spectrum contains a wealth of information on the intergalactic medium.

6.3 Ingredients for Generating Quasar Absorption Spectra

There are four quantities that go into the computation of the optical depth: temperature, peculiar velocity, overdensity and neutral fraction. That the temperature and peculiar velocity are important should be obvious from the expression for the absorption cross-section in equation (6.5). The temperature determines the extent of thermal broadening and the peculiar velocity changes the frequency of the photon in the fluid rest-frame (equation [6.3]). Let us define carefully what we mean by the other two quantities, the overdensity and the neutral fraction.

Suppose $n_{\rm H}(\mathbf{x})$ is the total proper number density of all hydrogen species at position $\mathbf{x}$ and $\bar{n}_{\rm H}$ is its spatial average. The overdensity $\delta_{\rm b}$, which describes the variation in space of $n_{\rm H}(\mathbf{x})$, satisfies:

$$n_{\rm H}(\mathbf{x}) = \bar{n}_{\rm H}\left[1+\delta_{\rm b}(\mathbf{x})\right] \ , \ \rho_b(\mathbf{x}) = \bar{\rho}_b\left[1+\delta_{\rm b}(\mathbf{x})\right] . \tag{6.8}$$

In the first expression, $\delta_{\rm b}$ is defined as the number overdensity of hydrogen. In the second expression , we equate $\delta_{\rm b}$ with the mass overdensity of baryons (ρ_b is the proper mass density of baryons and $\bar{\rho}_b$ is its mean), which is an excellent approximation for our application because there is no significant conversion of hydrogen into other elements, nor is there any interaction that could cause the spatial distribution of hydrogen to deviate significantly from that of other types of baryons.

What the Lyman-α absorption directly probes is not the total hydrogen density but its neutral component. The neutral fraction $X_{\rm HI}$ is defined by the following relation:

$$n_{\rm HI}(\mathbf{x}) = n_{\rm H}(\mathbf{x}) X_{\rm HI}(\mathbf{x}) \,, \tag{6.9}$$

where $n_{\rm HI}$ is the proper number density of neutral hydrogen as a function of space. The neutral fraction is determined by the balance between recombination and ionization, the rates of which are dictated by the temperature and radiation intensity respectively.

In general, all four quantities, overdensity $\delta_{\rm b}$, peculiar velocity $v_{\rm pec}$, temperature T and neutral fraction $X_{\rm HI}$, are functions of position. In the next two subsections, we discuss first how to determine the spatial distributions of $\delta_{\rm b}$ and $v_{\rm pec}$, and second how T and $X_{\rm HI}$ vary with position through their dependence on $\delta_{\rm b}$. All quantities are evaluated at the redshift $\bar{z}$ (see equation [6.3]).

6.3.1 The Zel'dovich Approximation

In cosmological models where dark matter (a term we use interchangeably with non-interacting matter) dominates the mass density of the universe, $\delta_{\rm b}$ as defined in equation (6.8) coincides with the dark matter overdensity $\delta_{\rm DM}$ on large scales. $\delta_{\rm DM}$ is defined in an analogous manner as before (equation [6.8]):

$$\rho_{DM}(\mathbf{x}) = \bar{\rho}_{DM}[1 + \delta_{\rm DM}(\mathbf{x})] \,, \tag{6.10}$$

where ρ_{DM} is the mass density of dark matter at position $\mathbf{x}$ and $\bar{\rho}_{DM}$ is its mean. The equality $\delta_{\rm b} = \delta_{\rm DM}$ is merely a statement that the hydrogen density (or in other words, the baryon density, which we assume are proportional to each other) varies

with position in the same manner as the dark matter density does. This is true on large scales where gas pressure is insignificant compared to the gravitational attraction of the dark matter, provided the baryons and dark matter start out having the same spatial distribution, which is true for most popular cosmological models. Moreover, with no significant interaction that distinguishes between the two on large scales, the baryons and dark matter share the same peculiar velocity field. On small scales gas pressure can cause the spatial distributions of baryons and dark matter to differ. We will return to this point below.

Hence, on sufficiently large scales (how large is large, an obviously important question, will be addressed later), it is adequate to know the overdensity and peculiar velocity of the dark matter. The Zel'dovich approximation (Zel'dovich 1970) is a well-tested approximation to compute the density and velocity distributions of dark matter in the mildly nonlinear regime ($\delta_{\rm DM} \lesssim 5$) before orbit-crossing. The reader is referred to the article by Shandarin and Zel'dovich (1989) for a comprehensive review (see also Hui and Bertschinger 1996 for an alternative formulation of the approximation). For completeness, we include a brief discussion here.

The starting point of the Zel'dovich approximation is the following equation for the displacement of a given mass element or particle:

$$\mathbf{x}(\mathbf{q}, t) = \mathbf{q} + D_+(t)\nabla_{\mathbf{q}}\psi(\mathbf{q}), \tag{6.11}$$

The coordinate $\mathbf{q}$ is the initial position of the mass element and $\mathbf{x}$ is its comoving position as a function of time. The displacement is then $D_+(t)\nabla_{\mathbf{q}}\psi(\mathbf{q})$. Its time dependent part $D_+(t)$ is the linear growth factor (Peebles 1980), which, for a universe with critical matter density, can be equated with a, the Hubble scale factor. The time independent function $\nabla_{\mathbf{q}}\psi(\mathbf{q})$ is determined by initial conditions. Growing mode initial conditions dictate that it is curl-free and hence its form as the gradient of the potential ψ ($\nabla_{\mathbf{q}}$ is the gradient in q space).

Expressions for the peculiar velocity and overdensity follow immediately from

equation (6.11):

$$\mathbf{v}_{\rm pec} = a\dot{D}_+\nabla_{\mathbf{q}}\psi \ , \ 1+\delta_{\rm DM} = \det{}^{-1}\left[\delta_{ij} + D_+(t)\frac{\partial^2\psi}{\partial q_i \partial q_j}\right] . \tag{6.12}$$

The dot in the first expression denotes differentiation with respect to proper time t. The peculiar velocity is defined by $\mathbf{v}_{\rm pec} = a d\mathbf{x}/dt$. The second expression follows from mass conservation i.e. $(1+\delta_{\rm DM})d^3x = d^3q$.

The function $\psi(\mathbf{q})$ contains all the information on the specific cosmological model one chooses to study. It is related to the commonly used power spectrum P in the following way. Suppose $\delta_{\rm lin}(\mathbf{q})$ is the linear overdensity today. The Fourier counterparts of both quantities, denoted by the tilde, are related by:

$$\tilde{\psi}(\mathbf{k}) = -k^{-2}\tilde{\delta}_{\rm lin}(\mathbf{k}) \tag{6.13}$$

where k is the magnitude of the vector $\mathbf{k}$, the wavenumber. The power spectrum P is defined in terms of the two-point correlation of $\delta_{\rm lin}$:

$$\langle \delta_{\rm lin}(\mathbf{k})\delta^*_{\rm lin}(\mathbf{k}')\rangle \equiv P(k)\delta^3(\mathbf{k}-\mathbf{k}') \tag{6.14}$$

where the $*$ denotes complex conjugation. Using the following convention for the Fourier transform:

$$\delta_{\rm lin}(\mathbf{q}) = \int \tilde{\delta}_{\rm lin}(\mathbf{k})e^{i\mathbf{k}\cdot\mathbf{q}}d^3k \, , \tag{6.15}$$

it can be shown that the root-mean-squared linear overdensity fluctuation at a given time t is:

$$D_+^2(t)\langle\delta^2_{\rm lin}(\mathbf{q})\rangle = D_+^2(t)\int_0^\infty 4\pi P(k)k^2dk \, . \tag{6.16}$$

We adopt the convention that $D_+ = 1$ today. For $\Omega_0 = 1$ where D_+ can be equated to the Hubble scale factor, D_+^2 above can be replaced by $(1+z)^{-2}$. The rms linear overdensity fluctuation will turn out to be very important for our latter discussion.

To produce a realization of a given cosmological model, the procedure goes as follows: first, we use the corresponding power spectrum to generate the Gaussian

random field $\psi(\mathbf{q})$ on a grid; second, we displace particles from their grid positions (which are denoted by $\mathbf{q}$ above) according to equation (6.11) at the time t, which corresponds to some mean redshift $\bar{z}$ of interest (equation [6.3]); third, a peculiar velocity is assigned to each particle according to equation (6.12); finally, we use the TSC (Triangular-Shaped density Cloud) scheme (Hockney & Eastwood 1988) to interpolate the particle positions and velocities to become momentum and mass densities on the grid and divide one by the other to obtain the velocity itself. The interpolation to obtain mass density is our way of enforcing mass conservation, as is expressed in equation (6.12). In the last procedure, we smooth the momentum and mass density fields over a small number of grid cells (in fact, we use one and have checked that the precise number is not important as long as it is small) before performing the division to obtain the velocity field so that we have well-defined velocities even in places with zero density after the TSC interpolation (Kofman et al. 1994). Any line of sight can then be chosen through the simulation box and the above procedure gives the overdensity and peculiar velocity (in fact only the component parallel to the line of sight is needed) at each grid point on it.

The above procedure is very fast because there is no need to solve the Poisson equation. Nor is it necessary to integrate any equation of motion. One simply multiply the displacement of each particle by an appropriate factor of $D_+(t)$.

There are, however, two remaining issues we need to address if we are to apply the Zel'dovich approximation intelligently.

The Zel'dovich approximation yields predictions for $\delta_{\rm DM}$, which can be equated with $\delta_{\rm b}$ only on large scales. On smaller scales, two problems arise. Gas pressure exerts its influence on baryons but not on dark matter, at least not directly, causing $\delta_{\rm b}$ and $\delta_{\rm DM}$ to differ. Furthermore, orbit-crossing occurs on small scales, in which case the Zel'dovich approximation is known to break down because the post-collapse gravitational force is not treated properly (Kofman et al. 1990; Matarrese et al. 1992; Brainerd et al. 1993; Bagla & Padmanabhan 1994; Coles et al. 1993).

Jeans Scale Smoothing

The effect of gas pressure is to smooth the baryon density field compared to its dark matter counterpart. The length scale below which this becomes important is the Jeans scale. In linear theory, the baryon overdensity obeys the following equation in a dark matter dominated universe (Bi et al. 1992; Peebles 1993):

$$\frac{\partial^2 \tilde{\delta}_{\rm b}}{\partial t^2} + 2H\frac{\partial \tilde{\delta}_{\rm b}}{\partial t} = -4\pi G\bar{\rho}_{DM}\tilde{\delta}_{\rm DM} + \frac{\gamma k_B \bar{T}}{\mu a^2}k^2\tilde{\delta}_{\rm b} \tag{6.17}$$

where the tilde denotes functions in Fourier space as before, H is the Hubble constant, k_B is the Boltzmann constant, G is the Newton constant, $\bar{\rho}_{DM}$ is the average dark matter mass density, $\bar{T}$ is the average temperature of the gas and μ is the mean mass of each gas particle (for a fully ionized gas composed of hydrogen and helium with primordial abundances, it is about 0.6 times the proton mass). The relation between the temperature (not its average but its actual value) and $1+\delta_{\rm b}$ is described by γ, the temperature being proportional to $(1+\delta_{\rm b})^{\gamma-1}$.

The Jeans scale is defined by

$$k_{\rm J}^{-1} = \sqrt{\frac{\gamma k_B \bar{T}}{4\pi a^2 G\mu\bar{\rho}_{DM}}}\,. \tag{6.18}$$

For a dark matter dominated universe, one can replace $\bar{\rho}_{DM}$ by the total mean density of the universe. For $\gamma = 1.5$, $\bar{z} = 3$, $\bar{T} = 10^4$K and a flat universe, $k_{\rm J} = 16.8\,h_{100}\,{\rm Mpc}^{-1}$. Note that γ does not necessarily equal 5/3 unless the gas is fully ionized *and* behaves adiabatically.

For the special case of $\bar{T} \propto a^{-1}$, making use of an equation for $\tilde{\delta}_{\rm DM}$ which is similar to equation (6.17) except for the absence of the temperature term, it can be shown that

$$\tilde{\delta}_{\rm b}(\mathbf{k}) = \frac{\tilde{\delta}_{\rm DM}(\mathbf{k})}{1+k^2/k_{\rm J}^2}\,, \tag{6.19}$$

if one ignores decaying modes. It expresses in a quantitative way the expectation that the overdensity in baryons is the same as that of dark matter on large scales (low k) but is lower on small scales (high k). For $\bar{T}$ with other time dependence, solutions

for equation (6.17) are more complicated but the low and high k limits are the same: $\tilde{\delta}_{\rm b} = \tilde{\delta}_{\rm DM}$ for small k and $\tilde{\delta}_{\rm b} = \tilde{\delta}_{\rm DM} k_{\rm J}^2/k^2$ for large k (Bi et al. 1992).

We should emphasize, however, that equation (6.17) holds only in linear theory. To incorporate pressure effects in the mildly nonlinear regime, one can multiply $P(k)$ in equation (6.16) by a factor of $(1 + k^2/k_{\rm J}^2)^{-2}$ and use the resulting smoothed power spectrum to generate the Zel'dovich displacement field. See Reisenegger & Miralda-Escudé (1995) for an application of the same method in a slightly different context.

In practice, we use instead the Gaussian smoothing kernel $e^{-k^2/k_{\rm J}^2}$ and have checked that both ways of smoothing produce very similar column density distributions in our simulations for the same initial $P(k)$. The Gaussian kernel is convenient because we use the same form of smoothing for nonlinear scale truncation, our next subject.

Nonlinear Scale Truncation

There exists a well-known cure to the problem of orbit-crossing which has been extensively tested (Coles et al. 1993; Melott et al. 1995). The basic idea is to smooth the initial power spectrum on small scales so that the amount of orbit-crossing that might have occurred is not significant enough to destroy the accuracy of the Zel'dovich approximation on large scales.

The initial power spectrum $P(k)$ is multiplied by a Gaussian smoothing kernel of the form $e^{-k^2/k_{\rm S}^2}$, which is then used to generate the Zel'dovich displacement field. The smoothing wavenumber $k_{\rm S}$ is chosen to be proportional to $k_{\rm NL}$ which defined as follows:

$$1 = D_+^2(t) \int_0^{k_{\rm NL}} 4\pi P(k) k^2 dk \,. \tag{6.20}$$

Note that $P(k)$ above is the initial power spectrum before any smoothing.

The proportionality constant between $k_{\rm S}$ and $k_{\rm NL}$ depends somewhat on the power spectrum, with more smoothing (smaller $k_{\rm S}$) required for those that have relatively more power on small scales (Melott 1994). We adopt

$$k_{\rm S} = 1.5\, k_{\rm NL} \,, \tag{6.21}$$

which has been shown to give good agreement with N-body simulations for CDM models (Melott et al. 1995).

The approximation employed in this way is known as the truncated Zel'dovich approximation.

For the cosmological models we consider in this paper, k_S ranges from about 2.3 Mpc^{-1} for some CDM models to more than 10 Mpc^{-1} for certain CHDM models. We compare k_S with k_J, the Jeans wavenumber, and choose the smaller of the two with which to smooth the initial power spectrum. The smoothed power spectrum is used to generate the Zel'dovich displacement field and the overdensity that follows from mass conservation is taken to be a good approximation to δ_b, at least on scales of interest.

In practice, among all the cosmological models we examine, only for the $\Omega_\nu = 0.2$ CHDM models in Sec. 6.8 are k_S's larger than k_J's, which is a reflection of the fact that these CHDM models have comparatively little power on small scales. We will argue further, in Sec. 6.8, that for these CHDM models, the precise smoothing scale is not important for the column densities of interest. Hence, uncertainty in the Jeans smoothing scale due to uncertainties in the temperature and equation of state of the intergalactic medium is not a concern.

For all the other models considered in the present paper, we choose k_S according to the prescription of equation (6.21). One may reasonably question whether by applying the truncated Zel'dovich approximation, one loses a significant amount of small scale fluctuations that should exist in reality. We will show in Sec. 6.5, by comparing to an actual hydrodynamic simulation, that for the range of column densities we are interested in, around $10^{12.5} - 10^{14.5}\,\mathrm{cm}^{-2}$, any loss of structure on scales smaller than the truncation scale does not affect the column density distribution significantly, if the standard prescription in equation (6.21) is followed. We will discuss the effects of departure from it in Sec.6.6.

An implicit assumption underlying the whole procedure of using the Zel'dovich approximation with a power spectrum smoothed on the length scale of k_S^{-1} or k_J^{-1}, is that the fluctuations on smaller scales do not contribute significantly to the number

of absorption lines at our column densities of interest, about $10^{12.5} - 10^{14.5}$ cm^{-2}. In particular, we have ignored the orbit-crossing which can shock-heat the gas. A way to justify our procedure is to make a plot of temperature versus density like Fig. 2 in Weinberg et al. (1996), from which one can infer that shock-heating is not important for regions of low overdensities or underdensities. It then remains to show that our column densities of interest do correspond to this regime of mild or negative overdensities, which we will do in Sec. 6.5. Alternatively, one can view the agreement of our approximate computation with a full hydrodynamic simulation as a justification of our assumption (Sec. 6.5).

From now on, we will refer to the overdensity predicted by the Zel'dovich approximation using the smoothed power spectrum as $\delta_{\rm b}$.

6.3.2 The Thermal and Ionization State

The evolution of temperature is governed by:

$$\frac{dT}{dt} = -2HT + \frac{2T}{3(1+\delta_{\rm b})}\frac{d\delta_{\rm b}}{dt} - \frac{T}{\sum_i q_i}\frac{d\sum_i q_i}{dt} + \frac{2}{3k_B n_b}\frac{dQ}{dt}, \qquad (6.22)$$

where d/dt is the Lagrangian derivative following each fluid element, n_b is the proper number density of all gas particles and T is the temperature which depends on both space and time. The symbol q_i is defined by $n_i \equiv (1+\delta_{\rm b}) q_i \bar{\rho}_b / m_p$, where n_i is the proper number density of the specie i, $\bar{\rho}_b$ is the mean mass density of baryons at the time of interest, m_p is the mass of the proton and $\delta_{\rm b}$ is the overdensity as in equation (6.8). For instance the mean number density of neutral hydrogen ($\bar{n}_{\rm H}$ in equation [6.8]) is $q_{\rm HI}\bar{\rho}_b/m_p$. The neutral fraction of hydrogen, $X_{\rm HI}$ as in equation (6.9), is then $q_{\rm HI}/(q_{\rm HI} + q_{HII})$. Note that q_i is a function of space and time in general.

The first two terms on the right hand side takes care of adiabatic cooling or heating. The third accounts for the change of internal energy due to the change in the number of particles. The last term dQ/dt is the heat gain (or negative heat loss) per unit volume by the gas particles from the surrounding radiation field. At a redshift of 2 to 4 and for densities of our interest, the main source of heat gain is photoionization

and the main source of heat loss is through the recombination of ionized hydrogen and the free electron. At higher redshifts, other processes become important, such as Compton cooling. More discussion on these processes will be presented in Hui and Gnedin (1996).

The above equation has to be supplemented by one that determines the abundance of each specie, which takes the form:

$$\frac{dq_i}{dt} = -q_i P + \sum_{j,k} q_j q_k n_b R \,. \tag{6.23}$$

For instance, if $q_i = q_{\rm HI}$, P is the the photoionization rate. It is given by:

$$P = \int_{\nu_{\rm HI}}^{\infty} 4\pi J_\nu \sigma_{\rm HI} \frac{d\nu}{h\nu} \,, \tag{6.24}$$

where h is the Planck constant, $h\nu_{\rm HI} = 13.6$ eV, $\sigma_{\rm HI}$ is the cross-section for photoionization as a function of the frequency ν and J_ν is the specific intensity. The photoionization rate P depends on the normalization as well as spectrum of J_ν. The specific intensity J_ν is generally taken to have a power law spectrum, ν^{-1} to $\nu^{-1.5}$, for frequencies just above $\nu_{\rm HI}$. The spectrum at higher frequencies is less important for the photoionization rate of hydrogen. A convenient way to hide our ignorance of the spectrum is to define $J_{\rm HI}$ (analogous to the definition in Miralda-Escudé et al. (1995) but differs by a factor of $10^{-21}\,{\rm ergsHz^{-1}s^{-1}cm^{-2}ster^{-1}}$):

$$J_{\rm HI} = \frac{\int_{\nu_{\rm HI}}^{\infty} 4\pi J_\nu \sigma_{\rm HI} \frac{d\nu}{h\nu}}{\int_{\nu_{\rm HI}}^{\infty} 4\pi \sigma_{\rm HI} \frac{d\nu}{h\nu}} (10^{-21}\,{\rm ergsHz^{-1}s^{-1}cm^{-2}ster^{-1}})^{-1} \,. \tag{6.25}$$

Then $P \sim 4 \times 10^{-12} J_{\rm HI}\, s^{-1}$. Note that $J_{\rm HI}$ is dimensionless.

Observations indicate that $J_{\rm HI}$ is between about 0.1 and 2.0 for $z = 2 - 4$ (Bechtold 1994; Batjlik et al. 1988; Lu et al. 1991b). A perhaps more common way of characterizing the radiation intensity is to quote its value, often referred to as J_{912}, at $\nu = \nu_\alpha$ or at wavelength 912Å, in units of ${\rm ergsHz^{-1}s^{-1}cm^{-2}ster^{-1}}$. The relation between J_{912} and $J_{\rm HI}$ depends on the spectrum. A good approximation for reasonable slopes of the spectrum ($J_\nu \propto \nu^{-m}$ for m between 1 and 1.5) is $J_{\rm HI} = 0.7 J_{912}/10^{-21}$.

For $q_i = q_{\rm HI}$, R is the recombination rate of ionized hydrogen and the free electron ($q_j = q_e$ and $q_k = q_{HII}$ in equation [6.24]):

$$R \sim 4 \times 10^{-13} (\frac{T}{10^4 K})^{-0.7} \, {\rm cm^3 s^{-1}} \, . \tag{6.26}$$

For $J_{\rm HI}$ with the values noted above, the photoionization time-scale is much shorter than the Hubble time. This means that hydrogen is essentially in ionization equilibrium. The two terms on the right hand side of equation (6.23) almost balance each other. Therefore,

$$X_{\rm HI} \sim 1.6 \times 10^{-6} \left(\frac{T}{10^4 K}\right)^{-0.7} \left(\frac{\Omega_b h_{100}^2}{0.0125}\right) \left(\frac{J_{\rm HI}}{0.5}\right)^{-1} (1+\delta_{\rm b}) \left(\frac{1+\bar{z}}{4}\right)^3 , \tag{6.27}$$

where we have assumed primordial abundances and that both helium and hydrogen are highly ionized. Both T and $\delta_{\rm b}$ can be position dependent.

We now have all the equations in place to compute the thermal and ionization evolution. The overdensity $\delta_{\rm b}$ is evolved using the Zel'dovich approximation. [2] Its rate of growth is substituted into equation (6.22), which is solved together with equation (6.23). The initial conditions are as follows. The gas temperature T is equal to the cosmic microwave background temperature at $z = 100$ (maintained by Compton scattering) and evolves adiabatically after that until the universe is reionized by the UV background. Abundances are assumed to be primordial, which is consistent with observations so far for column densities less than about $10^{14.5} \, cm^{-2}$ (See Songaila and Cowie (1996). Cooling processes due to metals are not important for our densities of interest in any case). All species are neutral until reionization occurs. One can integrate equations (6.22) and (6.23) forward starting from any time between $z = 100$ and the beginning of reionization.

Details of the computation are given in Hui and Gnedin (1996). We list the relevant conclusions. First, ionization equilibrium (equation [6.27]) is maintained at

[2] Note that there is no need to generate a full three-dimensional realization for this purpose. It is simpler to generate a set of eigenvalues of the matrix $\partial^2 \psi / \partial q_i \partial q_j$ according to the prescription of Doroshkevich (1970). The density is then determined through equation (6.12).

high accuracy except during the period of initial reionization. Second,

$$T = T_0(1 + \delta_b)^{\gamma-1}\,, \tag{6.28}$$

where T_0 is not dependent on position, is an excellent approximation for overdensities of interest, $\delta_b \lesssim 5$, with a little flattening at the low end (δ_b close to 0.1) for some reionization scenario. Third, both T_0 and γ depend on the reionization history, the reasonable ranges being $1.2 < \gamma < 1.7$ and $3000\,K < T0 < 30000\,K$ at $z = 3$. Note that the range of γ includes the particular case when the last two terms on the right hand side of equation (6.22) are neglected, in which case the exact solution is $T \propto a^{-2}(1 + \delta_b)^{2/3}$, the result of adiabatic expansion or compression.

In conclusion to this section, we have outlined a procedure for obtaining δ_b, v_{pec}, T and X_{HI}, all of which enter into the calculation of the optical depth τ (equation [6.1] or [6.5]). We can compute $e^{-\tau}$, called the transmission, which is the ratio of the emitted to the observed intensities. Observationally, its measurement requires knowledge of the quasar emission spectrum. Moreover, one must carefully choose the range of frequencies to consider if one is to limit the source of absorption to that due to the Lyman-α transition. For a discussion of these issues, the reader is referred to Press, Rybicki and Schneider (1993). To produce a realistic spectrum, one should also add noise and convolve the transmission with a window function to mimic instrumental resolution. This is important for a detailed comparison between theories and observations, which we will defer to latter work. Our x space grid cells, depending on which particular simulation, have sizes ranging from $0.028 - 0.05\,\mathrm{Mpc}$. Note that the true resolution in velocity space is not uniform because peculiar velocity varies from one place to another. Without peculiar velocity, the above grid cell sizes correspond to velocity cells of $2.8 - 5\ \mathrm{kms}^{-1}$, for $h_{100} = 0.5$ at $z = 3$ (equation [6.3]). The true velocity resolution is probably a little worse than that. As a comparison, high quality Keck Telescope data have a Full-Width-Half-Maximum of about $7\,\mathrm{km\,s}^{-1}$ and signal to noise per pixel of the order of 30 or higher (Hu et al. 1995; Lu et al. 1991a).

6.4 The Peculiar Velocity: its Effects on Line Shapes and the Column Density Distribution

We will show in this section that while the peculiar velocity plays an important role in determining the absorption profiles, its effect on the column density distribution is minimal. The procedures to obtain the column density distribution are discussed first.

6.4.1 Line Identification and the Column Density Distribution

Fig. 6-1 and Fig. 6-2 show the velocity, density and transmission ($e^{-\tau}$) along two lines of sight for a $\sigma_8 = 0.7$ CDM simulation, with $h_{100} = 0.5$. The significance of the dashed transmission profile will be explained in the next sub-section. The thermal and ionization parameters are described in the caption of Fig. 6-1. The nonlinear truncation scale k_S is 2.3 Mpc^{-1}. The transfer function is taken from Ma (1996). We find that using the BBKS transfer function (Bardeen et al. 1986) makes almost no difference to the resulting column density distribution, for the range of column densities we study.

The first thing to note is that for the given parameters,

$$b \equiv \sqrt{\frac{2k_BT}{m_p}} = 13 \text{ kms}^{-1}(1+\delta_b)^{1/4}. \tag{6.29}$$

This might seem to be too small because the observed lower limit of the b-value is about $15 - 20$ kms^{-1} (Hu et al. 1995; Lu et al. 1991a). A distinction should be made between the observed b-value and the b defined above. The observed b-value is obtained by fitting the quasar spectrum with superpositions of the Voigt profiles. Each Voigt profile yields a column density and a b-value. All the density peaks that give rise to absorption troughs in Fig. 6-1 and Fig. 6-2 have velocity widths larger than the small thermal width defined in equation (6.29). Therefore the narrow-maximum-

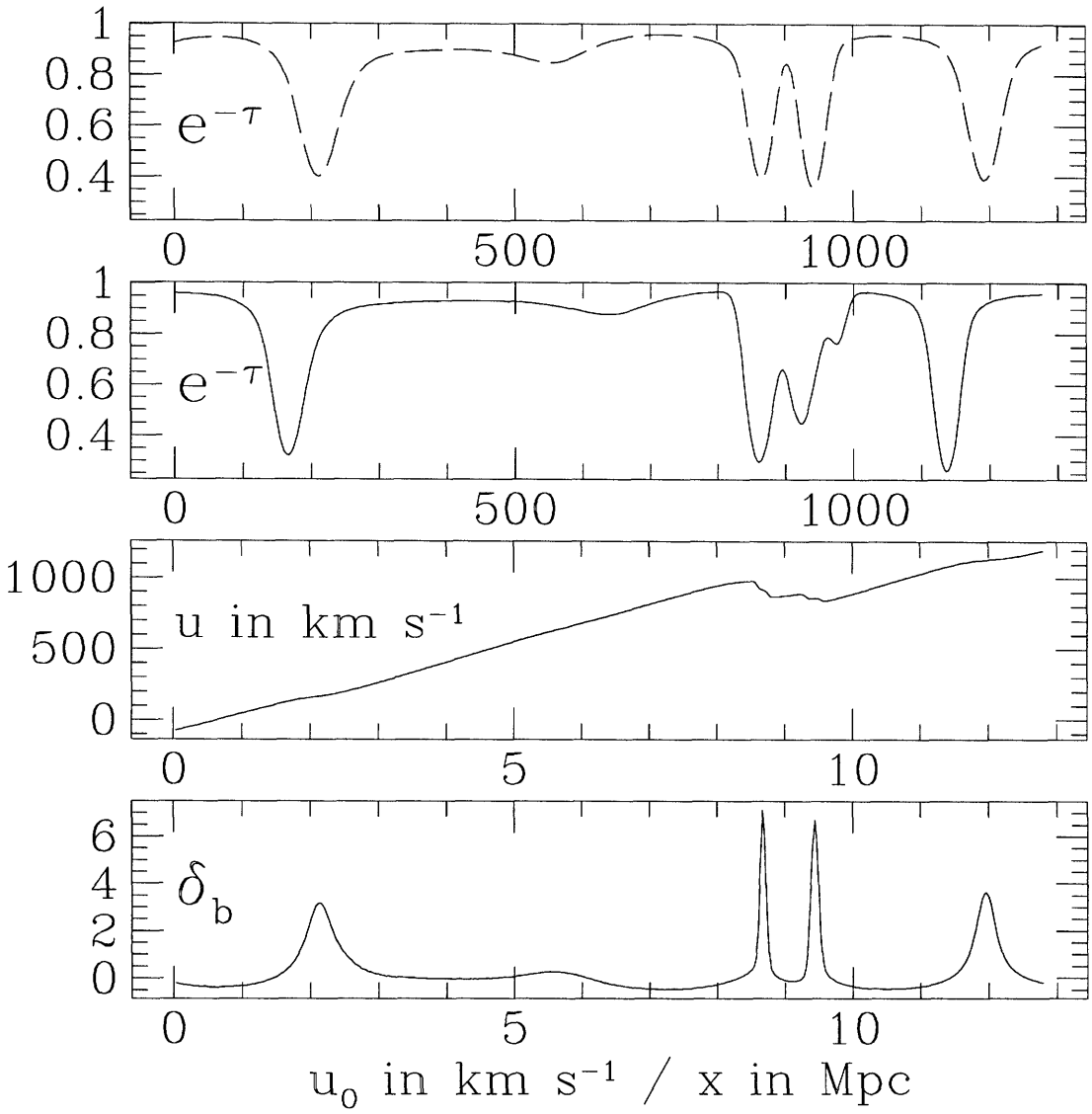

Fig. 6-1.— A line of sight through a $\sigma_8 = 0.7$ CDM simulation (produced using the truncated Zel'dovich approximation) at $\bar{z} = 3$, with $h = 0.5$. The transfer function is taken from Ma (1996) Box size is 12.8 Mpc with a grid spacing of 0.032 Mpc. The parameters are $\Omega_b h^2 = 0.0125$, $J_{\rm HI} = 0.5$, $T_0 = 10^4$ K ,$\gamma = 1.5$ and $k_{\rm S} = 2.3\,{\rm Mpc}^{-1}$. All distances are comoving. See Sec. 6.3 for definitions of the symbols. The abscissas for the lower two panels are the comoving distances along the line of sight in units of Mpc. The lower of the two panels is the profile of overdensity $\delta_{\rm b}$ (eq. [6.8]) and the upper one is the profile of velocity u (eq. [6.3]). The top two panels are both transmission profiles where τ is the Lyman-α optical depth and the abscissas represent u_O, which is related to the observed frequency through equation (6.4). The profile with solid line is obtained using the full density and peculiar velocity fields. The profile with dashed line is obtained using the density field but setting the peculiar velocity to zero everywhere (in which case, u becomes linear in x).

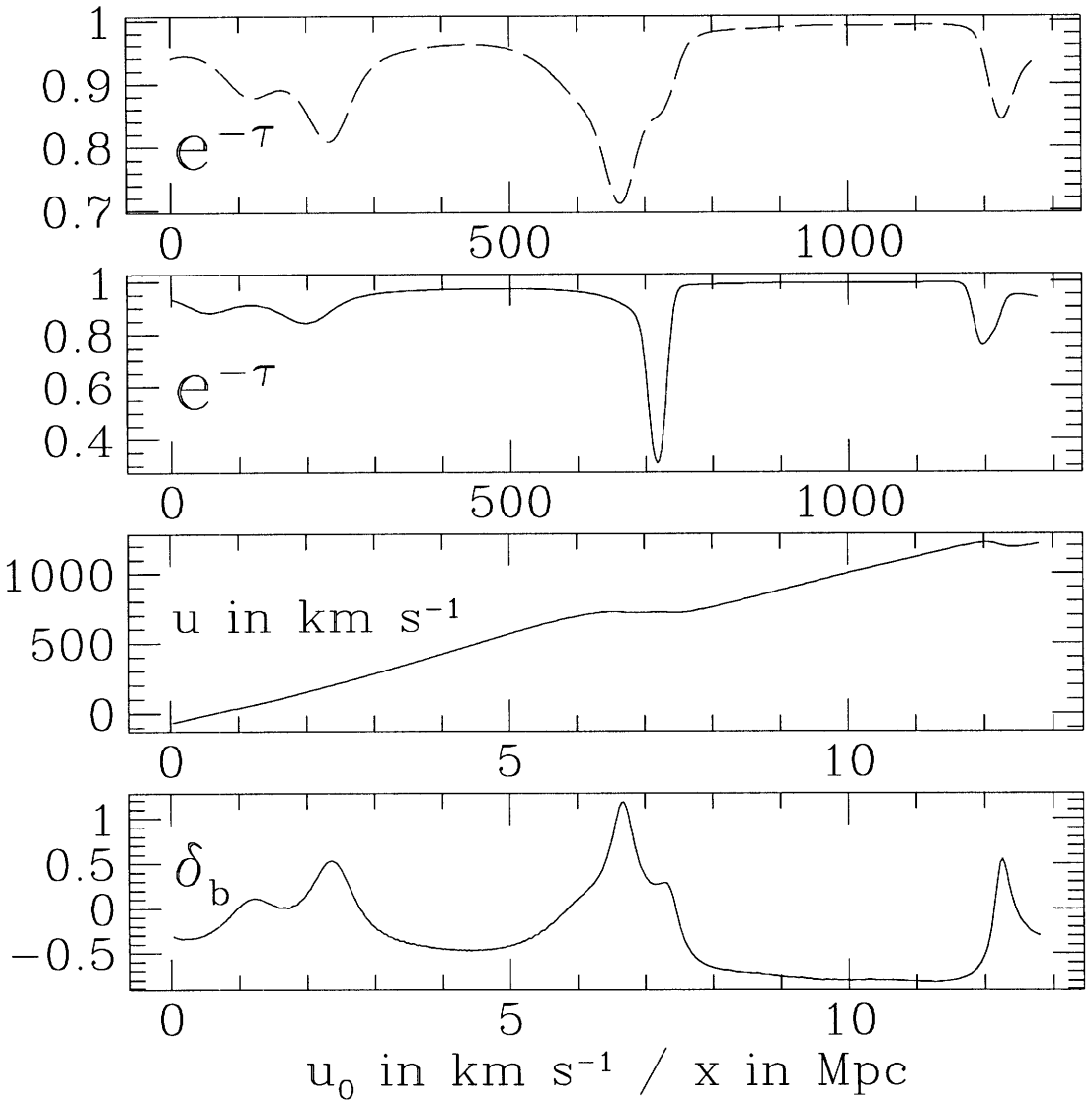

Fig. 6-2.— The same as in Fig. 6-1 but a different line of sight.

limit (equation [6.6]) does not apply and the absorption troughs do not exactly have the Voigt profile shapes. One can still search for the best-fit Voigt profile, taking into account the presence of noise, but the b-value obtained does not correspond to the thermal width in equation (6.29). It should also be emphasized that the recent hydrodynamic simulations of the Lyman-α forest, which have been so successful in accounting for a lot of its observed properties, have similarly low temperatures (see for instance Weinberg et al. 1996 [3])

One might wonder if there exists an alternative spectrum reduction method where the Voigt profile is not assumed to be the fundamental shape of absorption troughs, and for such a method, how the column density is assigned to each trough. The Voigt-profile-fitting-technique is nonetheless very important because it is how all existing observational data on the column density distribution are obtained.

An alternative line identification algorithm was proposed by Miralda-Escudé et al. (1995) and was also used by Hernquist et al. (1995). A transmission threshold is chosen, say 0.7. Any part of the spectrum that is continuously below the threshold is identified as an absorption line. The column density associated with it is defined by

$$N_{\rm HI} \equiv \frac{1}{\sigma_{\alpha 0}} \int_{\rm line} \tau(u_O) \frac{du_O}{c}\,, \tag{6.30}$$

where $\sigma_{\alpha 0}$ is defined after equation (6.5). The limits of integration are taken to be over the absorption line, i.e. where the transmission is continuously below the threshold. Note that if the narrow-maximum-limit or the thin cloud assumption were to hold, equation (6.6) can be substituted into equation (6.30) to show that $N_{\rm HI}$ does correspond to $\int n_{\rm HI} dx/(1+z)$, assuming the threshold is high enough so that most of the Voigt profile is included in the definition of the absorption line.

Let us call the above procedure the Threshold-Algorithm. We show in Fig. 6-3 the column density distribution computed according to the algorithm. The symbol

[3]Their output is at redshift of 2 and so naturally they have a lower temperature. In general, the temperature T_0 is dependent on the reionization history of the universe: crudely speaking, the closer the epoch of reionization is to the epoch of observation ($\bar{z} = 3$ in our case), the higher the temperature. Assuming reionization occurs before a redshift of 5, say, puts an upper bound on T_0 (Hui and Gnedin 1996).

$d^2N_{\mathrm{Ly}\alpha}/dN_{\mathrm{HI}}/dz$ denotes the number of absorption lines per unit redshift per unit column density. The reason for the chosen range of column densities will be given in Sec. 6.5. Two different transmission thresholds are used, 0.7 and 0.89 (Open triangles and crosses, ignoring the squares for the moment). The former is used by Hernquist et al. (1995) while the latter is the mean transmission of the spectrum. Changing the transmission threshold mainly affects the number of absorption lines with low column densities. The number of lines increases with a raised threshold. There is almost no effect, however, on the number of higher column density absorption lines.

As we will demonstrate later, the Threshold-Algorithm has the tendency to underestimate the number of absorption lines compared to the Voigt-profile-fitting-technique. One reason is that it does not deblend. In other words, a given absorption line according to the Threshold-Algorithm may contain more than one minimum in transmission. Such an absorption line would be broken up to a few lines if the Voigt-profile-fitting-technique is employed. To demonstrate this effect, we modify the Threshold-Algorithm: for each (parent) absorption line identified, we break it up into individual components (children) where each component is bordered by local maxima in the transmission within the confines of the parent. The column density for each child component is defined similarly as in equation (6.30) and the limits of integration are taken to be the boundaries of each component. We will call it the Threshold-Deblending-Algorithm.

The resulting column density distribution is denoted by the square symbols in Fig. 6-3 for the transmission threshold of 0.89. One can see that indeed the number of lines of low column densities go up. We should emphasize however that the Threshold-Deblending-Algorithm cannot be used to analyze observational data without modifications because in real life, noise creates local transmission maxima within any parent absorption line.[4] The question remains, though, how well the column density distribution computed using the Threshold-Algorithm, with or without de-

[4]In fact, numerical noise can also have the same effect. We check it by defining local maxima in two ways: local maxima over three cells and local maxima over five cells with the slope on either side of the maxima not changing signs. It turns out the resulting column density distributions are almost the same.

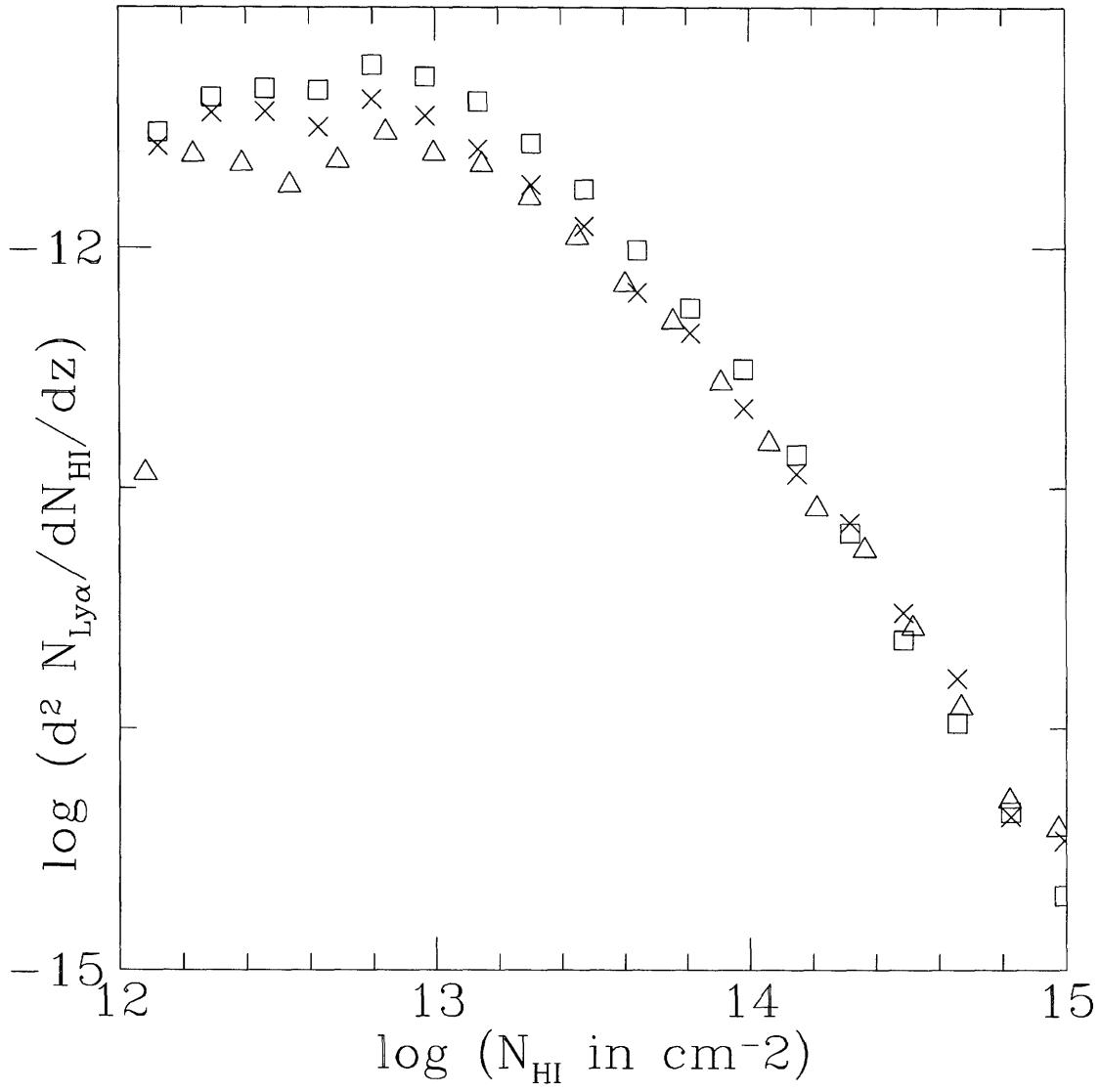

Fig. 6-3.— The column density distributions for the same model and parameters as those in Fig. 6-1. The quantity $d^2N_{\mathrm{Ly}\alpha}/dN_{\mathrm{HI}}/dz$ has units of cm^2. Crosses (transmission threshold of 0.89, which is also the mean transmission) and open triangles (transmission threshold of 0.7) represent the distributions computed using the Threshold-Algorithm. Open squares are obtained by applying the Threshold-Deblending-Algorithm at the threshold of 0.89.

blending, agrees with that obtained using the Voigt-profile-fitting-technique. This has been partially addressed by Miralda-Escudé et al. (1995). They applied the Threshold-Algorithm to spectra composed of randomly superposed Voigt profiles drawn from a known column density distribution and compared the measured distribution with the seed. They found good agreement for column densities larger than about 10^{14} to $10^{15}\,\mathrm{cm}^{-2}$ (Fig. 15[a] in their paper). We will address this question for lower column densities in Sec. 6.5.

For now, the Threshold-Algorithm is adopted as a simple way to identify lines and assign column densities, which we will use to study the effects of the peculiar velocity on the column density distribution.

6.4.2 The Role of Peculiar Velocities

The following experiment is performed to investigate the importance of peculiar velocities. We generate absorption spectra and compute the column density distribution using the same density field as that used to produce the solid curves in Fig. 6-1, Fig. 6-2 and the points in Fig. 6-3 but we set all peculiar velocities to zero.

Let us first examine some examples of the absorption spectra. The dashed curves in Fig. 6-1, Fig. 6-2 are the resulting spectra after putting all peculiar velocities to zero.

A comparison of the dashed absorption spectrum with its solid counterpart in each figure shows that the peculiar velocities play an important role in determining the shapes of absorption lines. Without peculiar velocities, the shapes of absorption troughs mirror closely those of the density peaks while with nonzero peculiar velocities, the absorption troughs can have quite different shapes from the underlying density field. Peculiar velocities can add or erase structures. An example of the former can be found in the pair of density peaks around $x = 9\,\mathrm{Mpc}$ and their corresponding absorption profiles in Fig. 6-1. An example of the latter can be found in the density peak(s) around $x = 7\,\mathrm{Mpc}$ and the corresponding absorption trough(s) in Fig. 6-2.

Broadly speaking, the effects of peculiar velocities on absorption spectra fall

into three categories. They are distinguished by the value of du/dx. First, there are regions in space where the peculiar velocity gradient is small so that du/dx is almost equal to its Hubble value $\bar{H}/(1+\bar{z})$ (equation [6.3]). An example is the density peak around $x = 2.2\,\mathrm{Mpc}$ in Fig. 6-2. The peculiar velocity shifts the position of the associated absorption trough but does not affect its shape.

Second, there are places where the peculiar velocity gradient is opposite in sign and comparable in magnitude to $\bar{H}/(1+\bar{z})$, in which case $|du/dx|$ becomes very small. Suppose also that $|d^2u/dx^2|$ is small. The implication is then a small range in u corresponds to a relatively large range in x. See for instance the density peak(s) around $x = 7\,\mathrm{Mpc}$ in Fig. 6-2, which is a really broad structure in x space but is relatively narrow in u space. The small $|du/dx|$ or the converging peculiar velocity flow around it helps produce a narrow absorption trough (second panel from the top in Fig. 6-2). Contrast it with the corresponding absorption feature in the top panel of the same figure, where peculiar velocities are set to zero. The limiting case where $|du/dx|$ exactly vanishes is called a velocity caustic (McGill 1990).

Third, there are regions where the peculiar velocity gradient dominates in such a way that du/dx is negative and $|du/dx|$ is not small. An example can be found around the pair of density peaks at $x = 9\,\mathrm{Mpc}$ in Fig. 6-1. This is where multiple-streaming occurs. A given range in u corresponds to disjoint pieces in x space. As a result, the shapes of the associated absorption troughs are significantly different from those of the underlying density peaks.

The three categories can be shown to correspond to the three evolutionary stages of a pancake collapsing along the line of sight. Restricting equations (6.11) and (6.12) to one dimension and putting it into equation (6.3), one obtains:

$$\frac{du}{dx} = \dot{a}\left(1 + 2a\frac{d^2\psi}{dq^2}\right)\left(1 + a\frac{d^2\psi}{dq^2}\right)^{-1}, \tag{6.31}$$

where we have assumed a universe of critical matter density so that $D_+ = a$. Re-

stricting equation (6.12) to one dimension, it can also be shown that

$$1 + \delta_{\rm DM} = \left(1 + a\frac{d^2\psi}{dq^2}\right)^{-1} , \tag{6.32}$$

where we have equated $\delta_{\rm DM}$ with $\delta_{\rm b}$ assuming the appropriate initial smoothing has been carried out as indicated in Sec. 6.3.1. As a grows, it can be seen that du/dx goes through the three different regimes outlined above. At the velocity caustic where $du/dx = 0$, it can be shown that $\delta_{\rm DM} = 1$ (McGill 1990). This conclusion does not hold in general of course because pancakes can collapse in directions different from the line of sight. But it is true that velocity caustics are often found in regions of slight overdensities.

In principle, at a velocity caustic, an absorption line can arise without even any variation in the density field at all (equation [6.5]). In practice, one expects that converging peculiar velocity flows are accompanied by density peaks. This is consistent with the few examples we have seen.

Next, we consider how the column density distribution changes when the same density field is used but all the peculiar velocities are put to zero. This is shown in Fig. 6-4.

The mean transmission of the analysis with zero peculiar velocities differ from the mean transmission of the full analysis by less than a percent. It is used as the transmission threshold in the line identification procedure for both analyzes. The resulting column density distributions are very similar.

Hence, the peculiar velocity plays a relatively minor role in determining the column density distribution. It changes the shapes of absorption troughs without altering the overall number of lines and their column densities. This serves to motivate an approximation we will introduce in the next section.

A final note on velocity caustics. The reader might worry that at a velocity caustic, the optical depth may diverge while it is clear from equation (6.1) that for a finite number density of neutral hydrogen, the optical depth should always be a finite quantity. The resolution is that close to a velocity caustic at $u = u_c$, du/dx

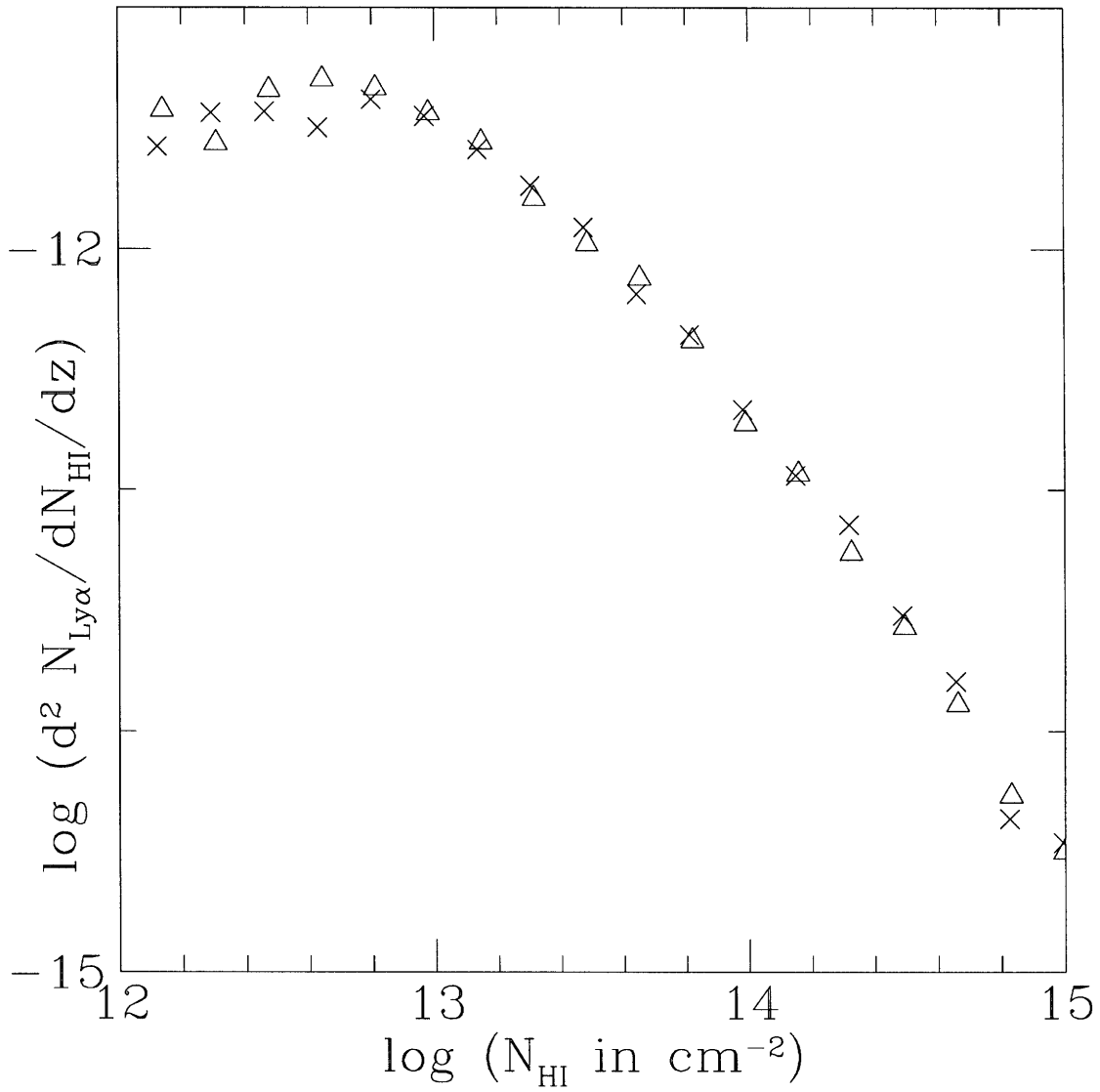

Fig. 6-4.— Same parameters as in Fig. 6-1. The column density distributions are computed using the Threshold-Algorithm. Crosses (the same as crosses in Fig. 6-3): nonzero peculiar velocities. Open triangles: peculiar velocities set to zero. The transmission threshold is 0.89.

goes like $(u-u_c)^{1/2}$ (provided the second derivative of u with respect to x is nonzero, otherwise it will be $(u-u_c)^{2/3}$ if the third derivative does not vanish and so on, by simple Taylor series expansion; see Shandarin and Zel'dovich 1989). So under the integration in equation (6.5), the optical depth remains finite. We note also that because of the singular nature of $(du/dx)^{-1}$ around $u = u_c$, the derivation leading to equation (6.7) breaks down at a velocity caustic.

6.5 The Statistics of Density Peaks

In this section we explore a simple approximation in which each density peak in x space is identified with an absorption line. This is motivated by the fact that peculiar velocities do not play a major role in determining the column density distribution and that each maximum in density corresponds to a minimum in the absorption spectrum if the peculiar velocities are set to zero and if the maximum in density are separated by a distance larger than that given by the thermal broadening width.

To calculate $d^2N_{\mathrm{Ly}\alpha}/dN_{\mathrm{HI}}/dz$, we relate dz and dx by ignoring peculiar velocities: $dz = c^{-1}\bar{H}dx$. Hence

$$\frac{d^2N_{\mathrm{Ly}\alpha}}{dN_{\mathrm{HI}}dz} = \frac{d^2N_{\mathrm{pk}}}{dN_{\mathrm{HI}}dx}\frac{c}{\bar{H}}, \tag{6.33}$$

where dN_{pk}/dx is the average comoving number density of density peaks along a random line of sight, $\bar{H}$ is the Hubble constant at the redshift of interest.

For each density peak, we need a simple prescription for assigning a column density. To that end, we do the following expansion around each density maximum:

$$\ln[n_{\mathrm{HI}}(x)] = \ln[n_{\mathrm{HI}}(x_{\mathrm{pk}})] + \frac{1}{2}\left.\frac{d^2\ln[n_{\mathrm{HI}}]}{dx^2}\right|_{x=x_{\mathrm{pk}}}(x-x_{\mathrm{pk}})^2. \tag{6.34}$$

It is a straightforward Taylor expansion around the position of the peak x_{pk}. The second derivative in the last term is negative. The rationale behind expanding $ln[n_{\mathrm{HI}}]$ rather than n_{HI} itself is that n_{HI} is supposed to fall off quickly far away from the peak (until, of course, another peak is encountered). In other words, the above

expansion implies that $n_{\rm HI}$ has a Gaussian fall-off (instead of a power-law one if $n_{\rm HI}$ itself were Taylor expanded). In a sense, this is close in spirit to the Voigt-profile-fitting-technique. Suppose the broad-maximum-limit (eq. [6.7]) holds so that the local optical depth is simply proportional to the number density of neutral hydrogen if one ignores peculiar velocities. Then, fitting a minimum in optical depth with the Voigt or thermal profile (eq. [6.5]) is equivalent to fitting the corresponding density peak with a Gaussian.

We then assign the following column density to the density peak:

$$N_{\rm HI} = \int_{\rm pk} \frac{dx}{1+\bar{z}} n_{\rm HI}(x) = n_{\rm HI}(1+\bar{z})^{-1} \sqrt{2\pi \left(\frac{-d^2 \ln[n_{\rm HI}]}{dx^2} \right)^{-1}} \Bigg|_{x=x_{\rm pk}} \tag{6.35}$$

where equation (6.34) has been used and where $\int_{\rm pk}$ denotes integration around the peak until it decays sufficiently. The above equation has been derived before by the authors using the Stationary Phase Method (Gnedin and Hui 1996).

Using equations (6.27) and (6.28), the above can be rewritten as:

$$N_{\rm HI} = 1.63 \times 10^{13} {\rm cm}^{-2} \left(\frac{T_0}{10^4 K} \right)^{-0.7} \left(\frac{\Omega_b h_{100}^2}{0.0125} \right)^2 \left(\frac{J_{\rm HI}}{0.5} \right)^{-1} \left(\frac{1+\bar{z}}{4} \right)^5 \left[\frac{2 - 0.7(\gamma - 1)}{1.65} \right]^{-0.5} A \tag{6.36}$$

where A is defined as

$$A \equiv (1+\delta_{\rm b})^{2-0.7(\gamma-1)} \left(\frac{-d^2 \ln[1+\delta_{\rm b}]}{dx^2} \right)^{-\frac{1}{2}} \Bigg|_{x=x_{\rm pk}}, \tag{6.37}$$

with x being measured in Mpc.

We will refer to our method as the Density-Peak-Ansätz. It consists of two parts: 1. associate each density peak in x space with an absorption line; 2. assign a column density to each density peak according to equation (6.36).[5]

Making use of equations (6.33) and (6.36), the column density distribution can

[5]Strictly speaking, care should be taken not to count peaks that are separated in velocity space by distance much less than the thermal width as contributing to more than one absorption line. We will address this later in the section.

be written as

$$\frac{d^2N_{\mathrm{Ly}\alpha}}{dN_{\mathrm{HI}}dz} = 6.25\times10^{-14}\,\mathrm{cm}^2 \left(\frac{T_0}{10^4K}\right)^{0.7} \left(\frac{\Omega_b h_{100}^2}{0.0125}\right)^{-2} \left(\frac{J_{\mathrm{HI}}}{0.5}\right) \left(\frac{1+\bar{z}}{4}\right)^{-5} \left[\frac{2-0.7(\gamma-1)}{1.65}\right]^{0.5} \frac{c}{\bar{H}} \frac{d^2N_{\mathrm{pk}}}{dAdx}, \tag{6.38}$$

where x is in units of Mpc. The computation of the last factor $d^2N_{\mathrm{pk}}/dA/dx$ involves counting the number density of peaks in x space having the quantity A within the range dA.

Let us define $\xi \equiv \ln[1+\delta_{\mathrm{b}}]$. Suppose one is given $P(\xi,\xi',\xi'')d\xi d\xi' d\xi''$ which is the probability that ξ and its first and second derivatives with respect to x fall in the specified ranges at a point. Then,

$$\frac{dN_{\mathrm{pk}}}{dx} = \int_{-\infty}^{\infty} d\xi \int_{-\infty}^{0} d\xi'' |\xi''| P(\xi,\xi'=0,\xi''), \tag{6.39}$$

where $\frac{dN_{\mathrm{pk}}}{dx}$ is the integral of $\frac{d^2N_{\mathrm{pk}}}{dAdx}$ over all A (Bardeen et al. 1986).

By a change of variable and a differentiation, one can obtain

$$\frac{d^2N_{\mathrm{pk}}}{dAdx} = \frac{1}{[2-0.7(\gamma-1)]A} \int_{-\infty}^{0} d\xi'' |\xi''| P(\xi,\xi'=0,\xi'') \tag{6.40}$$

where ξ should be expressed in terms of ξ'' and A using equation (6.37).

Note that the above two equations are completely general and no assumption about the Gaussianity of the underlying fields has been made. The hard part is of course to come up with the probability function P. The one point probability distribution of just ξ or density has been calculated for the Zel'dovich approximation (Kofman et al. 1994). We find the one point joint-probability distribution of density and its first and second derivatives difficult to calculate analytically for the Zel'dovich approximation, unless all particles are constrained to move in only one dimension, namely along the line of sight. A numerical approach is adopted in this paper and the number of peaks is counted along lines of sights in actual three-dimensional realizations. We will also discuss some analytical results based on the restrictive one-dimensional case. In a separate paper, we discuss an analytical calculation based upon not the Zel'dovich but the lognormal approximation, where ξ is assumed to be

a Gaussian random field (Gnedin & Hui 1996).

We test the Density-Peak-Ansätz in two different ways. First, we make a scatter plot of the column density obtained using the Threshold-Deblending-Algorithm versus the column density obtained by searching for the maximum density peak that contributes to each absorption line identified using the threshold method and then applying equation (6.36). The result is shown in Fig. 6-5. It shows that while the agreement is far from perfect, the column densities assigned using the Density-Peak-Ansätz and using the Threshold-Algorithm are broadly consistent.

The important question, however, is whether the Density-Peak-Ansätz, coupled with the Zel'dovich approximation, gives the correct number of absorption lines as a function of column density. We compare the column density distribution obtained using our approximate methods against that obtained by applying the Voigt-profile-fitting-technique to synthetic spectra from a full hydrodynamic simulation (Zhang et al. 1996). This is done in Fig. 6-6. We also put in the same figure the column density distribution obtained using the Threshold-Algorithm, coupled with the Zel'dovich approximation. The predictions of one-dimensional Zel'dovich approximation and the lognormal approximation are shown as well for comparison.

The level of agreement between the exact hydrodynamic computation and our calculation based on the Density-Peak-Ansätz coupled with the Zel'dovich approximation is encouraging. Two sets of points are shown for our approximate calculation using the Density-Peak-Ansätz, one (open triangles) with exactly the same box size and grid spacing as the hydrodynamic simulation and the other (open squares) with larger box size and smaller grid spacing. They both agree very well with the exact computation. We will explore the effects of changing the resolution in the next subsection. A third set of points (crosses) shows that the Threshold-Algorithm described in Sec. 6.4.2 underestimates the number of lines at low column densities.

We also show in Fig. 6-6 two sets of curves based on the lognormal approximation but using the same Density-Peak-Ansätz (see Gnedin and Hui 1996). One of them has the same amount of initial smoothing as that of the truncated Zel'dovich approximation and the other has less smoothing so as to match the final (not linear)

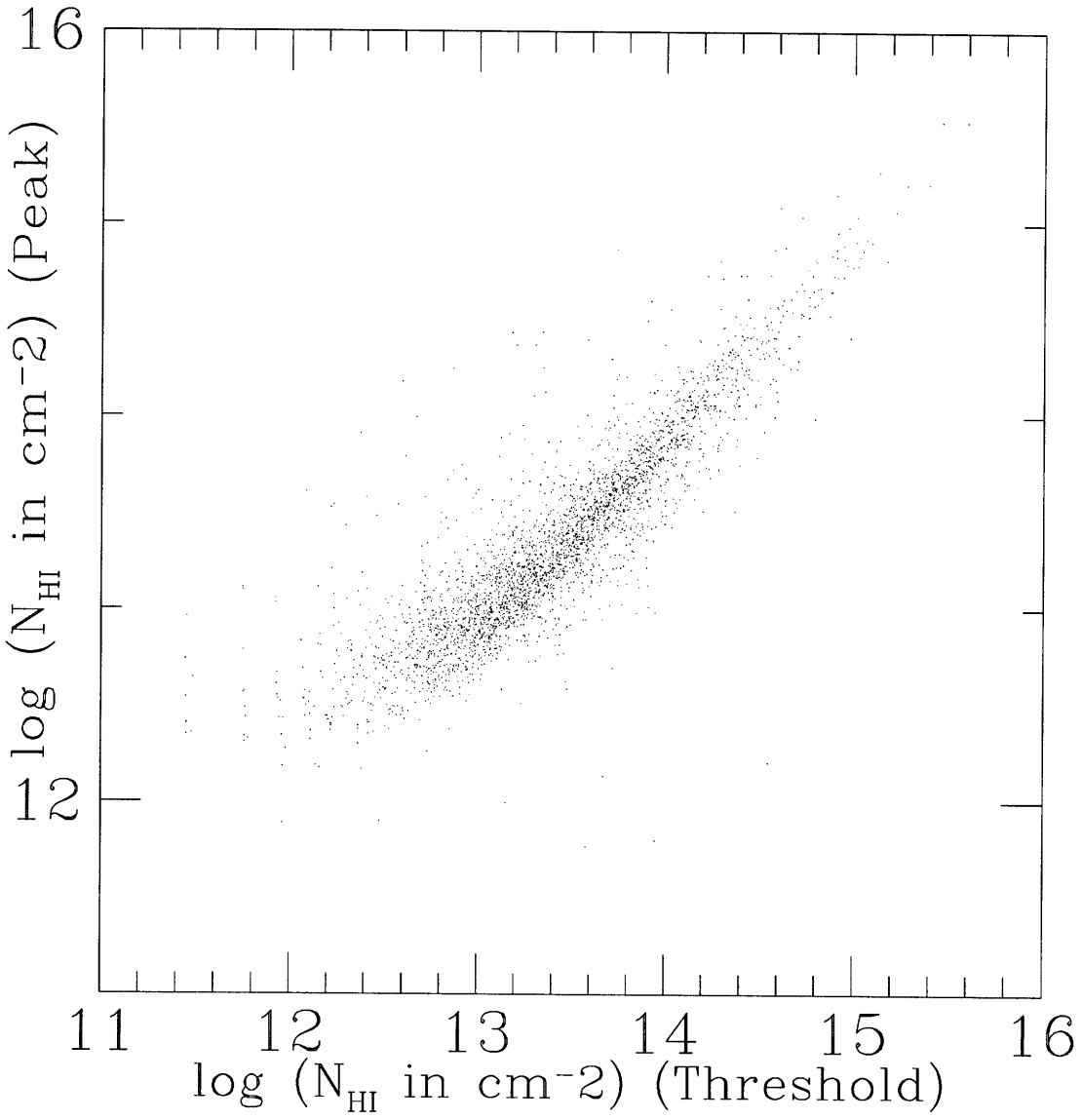

Fig. 6-5.— Same parameters as in Fig. 6-1. The column densities computed using two different methods are plotted against each other. First, we identify absorption lines by the Threshold-Deblending-Algorithm using a transmission threshold of 0.89 and assign column densities according to equation (6.30), which are plotted as the abscissas. We then take each absorption line identified using the Threshold-Deblending-Algorithm and search for the corresponding maximum in δ_b and apply equation (6.36) to assign a second set of column densities, which are plotted as the ordinates.

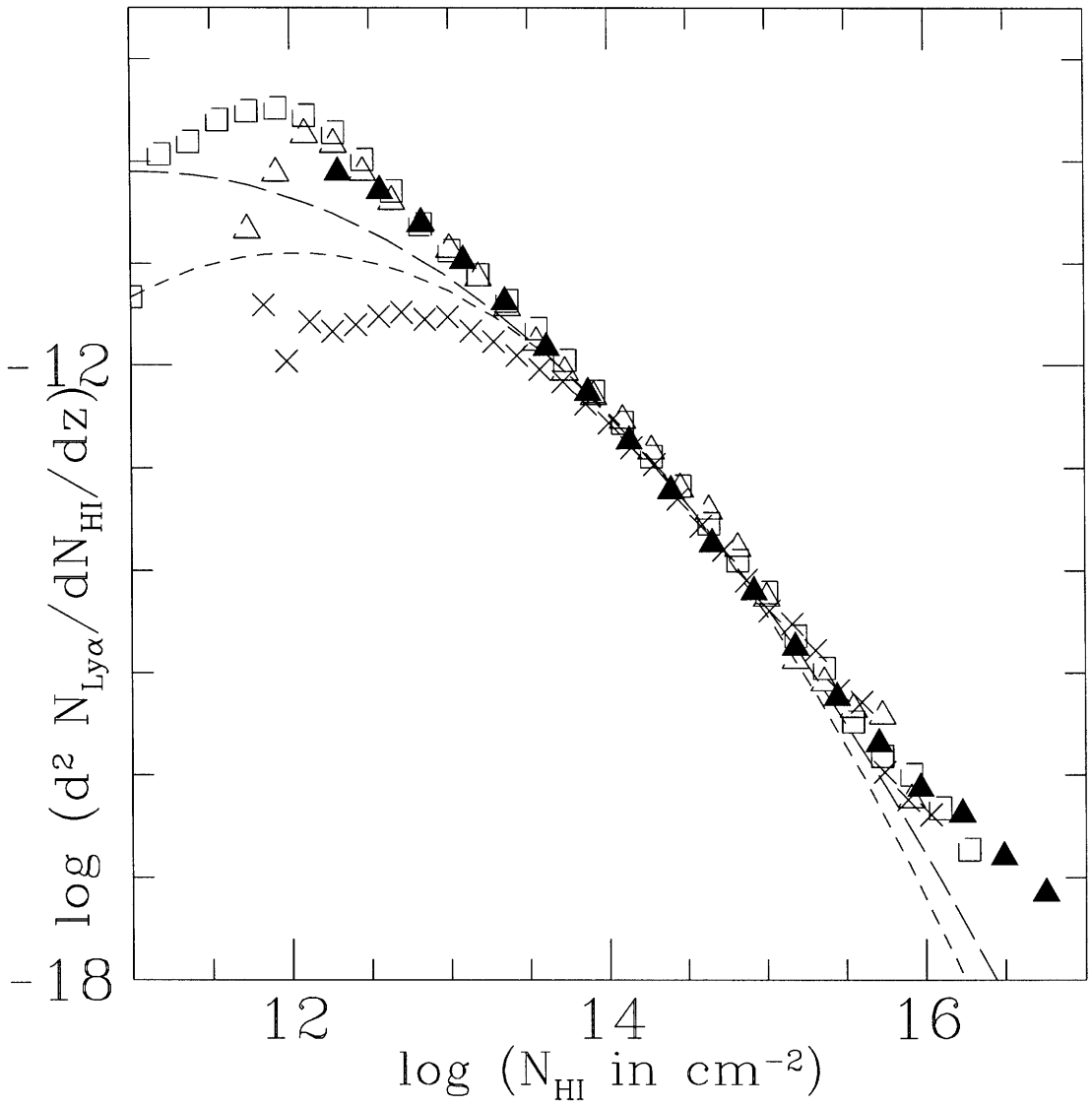

Fig. 6-6.— The model is $\sigma_8 = 0.7$ CDM at $\bar{z} = 3$ with $h = 0.5$, $\Omega_b = 0.06$, $J_{\rm HI} = 0.325$, $T_0 = 10^4 K$ and $\gamma = 1.45$. Solid triangles represent the distribution obtained by applying the Voigt-profile-fitting-technique to synthetic spectra from a full hydrodynamic simulation (Zhang et al. 1996) with box size of 9.6 Mpc comoving and grid spacing of 0.075 Mpc. Open triangles are the predictions of the Density-Peak-Ansätz coupled with the truncated ($k_{\rm S} = 2.3\,{\rm Mpc}^{-1}$) Zel'dovich approximation using the same box size and grid-spacing. Open squares and crosses are the results of applying the Density-Peak-Ansätz and the Threshold-Algorithm (transmission threshold at 0.83, the mean transmission) respectively to the output of the same truncated ($k_{\rm S} = 2.3\,{\rm Mpc}^{-1}$) Zel'dovich approximation with box size of 12.8 Mpc and grid spacing of 0.05 Mpc. The short-dashed and long-dashed curves are the predictions of the Density-Peak-Ansätz coupled with the lognormal approximation, the former with $k_{\rm S} = 2.3\,{\rm Mpc}^{-1}$ and the latter with the smoothing scale chosen so that the final (not linear) rms density fluctuation matches that of the Zel'dovich approximation ($k_{\rm S} = 3.6\,{\rm Mpc}^{-1}$).

rms density fluctuation of the Zel'dovich computation. Both underestimates the number of lines at low column densities. The lognormal approximation tends to predict too much flattening of the column density distribution at low column densities. (At the very low column densities, the lognormal approximation seems to give more lines than the Zel'dovich approximation but it is really a resolution effect: see the next sub-section.)

The reader might have noticed that we have included in Fig. 6-6 a wider range of column densities than is warranted by the nature of our approximations. For instance, objects with column densities higher than $10^{16}\,\mathrm{cm}^{-2}$ are almost certainly highly nonlinear and we do not expect the truncated Zel'dovich approximation to work well in this regime. For the low column densities, the finite resolution should cause us to underestimate the number of absorption lines. In the next sub-section, we give quantitative estimates of the range of column densities within which the Density-Peak-Ansätz, used in conjunction with the truncated Zel'dovich approximation, can be counted on to give reliable column density distributions.

We will also discuss two different ways of defining a density peak in the next sub-section: local maxima over three cells or local maxima over five cells with the slope on either side of the maxima not changing signs. The three-cell criterion is used in Fig. 6-6. One expects however that some of the three-cell peaks are not real but merely artifacts of numerical noise, especially those with low column densities. The five-cell criterion, on the other hand, probably fail to include some narrow peaks which are real. We will see that the two different criteria give almost identical results above a certain column density.

One aspect of the Density-Peak-Ansätz we have glossed over is that two density peaks that are separated by a distance in velocity space much less than the thermal width should not be counted as contributing to two but to one absorption lines. A more sophisticated approach would be to group together such density peaks and use the sum of their column densities as the column density of one single absorption line. We find that for the the range of validity discussed in the following sub-section, it makes little difference. It is conceivable, however, that this effect cannot be ignored

for simulations with higher resolution than what we have.

6.5.1 The Range of Validity

For the computation presented in Fig. 6-6, the column density (given by the Density-Peak-Ansätz) above which the mean δ_b exceeds 5 is about $10^{14.1}\,\mathrm{cm}^{-2}$. For the parameters listed in the caption of Fig. 6-6, $N_{\mathrm{HI}} = 3.6 \times 10^{13}\,A\ \mathrm{cm}^{-2}$ (equation [6.36]). We therefore take $A = 3.5$ as an upper limit beyond which we cannot expect our approximations to be reliable.

Note that according to Fig. 6-6, comparing with the hydrodynamic simulation data, the Density-Peak-Ansätz, together with the truncated Zel'dovich approximation, seems to give reliable number density of absorption lines for column densities higher than $10^{14.1} cm^{-2}$. The level of agreement at such high column densities (and by extension, such high δ_b) is surprising. We will take the conservative approach and adopt the $A = 3.5$ upper limit.

To illustrate the relation between column density and $1+\delta_b$, we make a log-log scatter plot of one versus the other in Fig. 6-7.

To determine the column density below which finite resolution results in an underestimate of the number of absorption lines, we perform a simulation using the truncated Zel'dovich approximation with the same parameters as in Fig. 6-6 but higher resolution: 12.8 Mpc box size and grid spacing of 0.0284 Mpc. A comparison of the resulting column density distributions is made in Fig. 6-8.

Note that we have included two definitions of density peaks (three-cell and five-cell). For each simulation, the true column density distribution is probably somewhere between the two in the places they differ.

We take the low column density cut-off to be $10^{12.8}\,\mathrm{cm}^{-2}$ for the lower resolution simulation (box size of 12.8 Mpc , with grid spacing of 0.05 Mpc) using the three-cell definition of peaks. It can be seen that the higher resolution simulation differs from the lower one only at column densities less than roughly this cur-off value. Moreover, above this column density, the three-cell and five-cell criteria give almost identical results.

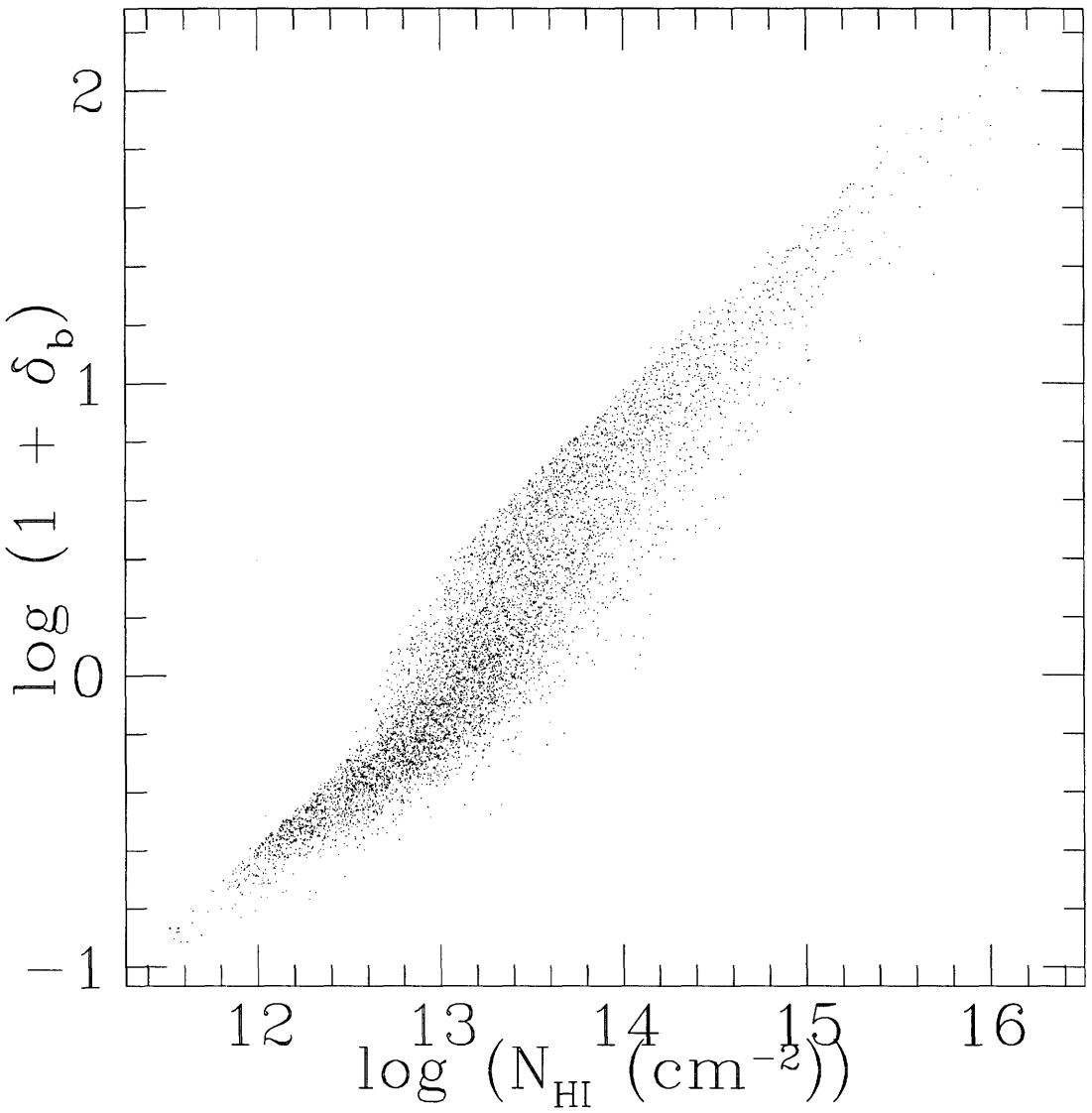

Fig. 6-7.— A log-log scatter plot of $1 + \delta_b$ at density peaks versus their column densities. The density field is obtained from the truncated Zel'dovich approximation using the same parameters as the open squares in Fig. 6-6. The column density is assigned using the Density-Peak-Anätz.

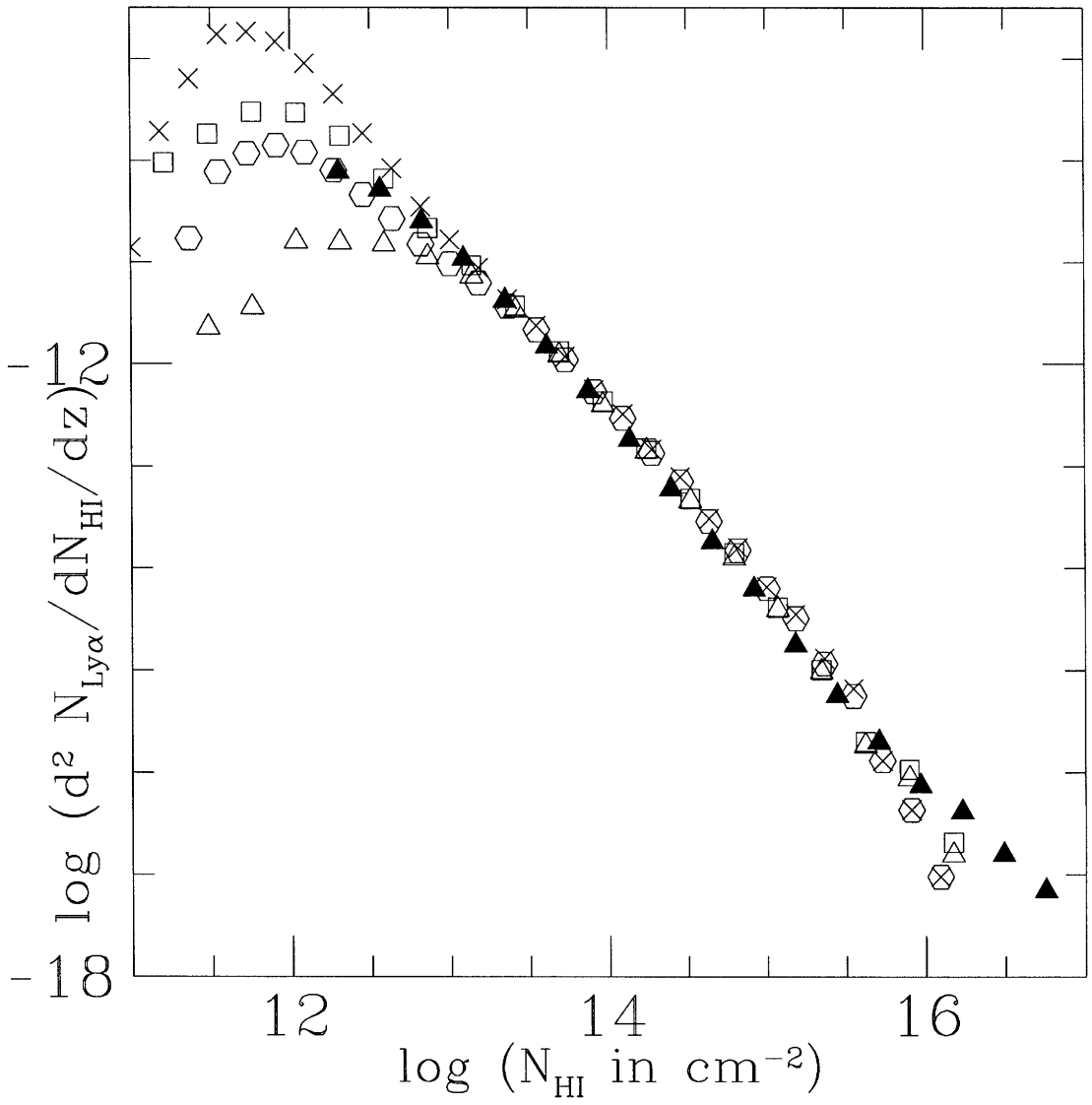

Fig. 6-8.— The column density distributions obtained using the Density-Peak-Anätz in conjunction with the truncated Zel'dovich approximation for two simulations of different resolutions but the same cosmological, thermal and ionization parameters as in Fig. 6-6. The open squares and open triangles represent the distributions of a simulation with box size of 12.8 Mpc and grid spacing of 0.05 Mpc (all distances quoted are comoving). The open squares here are the same as those in Fig. 6-6 where each maximum over three cells is identified as a peak. The open triangles are the result of a different definition of peaks: a local maximum over five cells with the density slope on each side of the maximum not changing signs. Similarly, the crosses and open hexagons are the distributions for a simulation of box size 12.8 Mpc and grid spacing 0.0284 Mpc, using the three-cell and five-cell definition of peaks respectively. The solid triangles are same as those in Fig. 6-6, representing the column density distribution obtained by applying the Voigt-profile-fitting-technique to synthetic spectra from a full hydrodynamic simulation.

The parameters in the Zel'dovich simulations in Fig. 6-8 are such that $N_{\rm HI} = 3.6 \times 10^{13}\, A\, {\rm cm}^{-2}$ (equation [6.36]). Hence the above cut-off implies a lower limit of 0.18 for A. From now on, we will use the three-cell definition of density peaks.

In the following section, we will systematically investigate how the column density distribution depends on the cosmological parameters and properties of the intergalactic medium. All the simulations presented in the next two sections have the same resolution and box size, 256^3 grid points with grid spacing of 0.05 Mpc. For each of them, we will only plot the part of the column density distribution that falls within the limits of $0.18 < A < 3.5$. The column densities these limits correspond to depend on the properties of the intergalactic medium and the redshift (equation [6.36]). Note how our conservative limits for A greatly reduce the range of column densities we can examine but within these limits, we can be reasonably confident that the truncated Zel'dovich approximation together with the Density-Peak-Ansätz should yield accurate predictions for the column density distribution.

6.6 The Column Density Distribution of the CDM models: Dependence on the Ionization Flux, Temperature, Equation of state and Cosmological Parameters

The Cold Dark Matter (CDM) models are used to study the influence of the cosmological parameters and properties of the intergalactic medium on the column density distribution. The tools we use to calculate the column density distributions are the truncated Zel'dovich approximation and the Density-Peak-Ansätz.

All CDM models discussed here have $\Omega_0 = 1$, $h_{100} = 0.5$ and are dark matter dominated. We use the Ma (1996) CDM transfer function. More detailed discussions of the simulations and the low and high column density cut-offs can be found at the end of Sec. 6.5.1. Table 6-1 contains a summary of all the CDM models discussed in this paper.

6.6.1 Dependence on Overall Temperature, Ionization Flux and Baryon Density

Let us first consider the CDM model with σ_8 and the spectral index n being fixed at 0.7 and 1 respectively. Let us also hold the equation of state at $T \propto (1+\delta_b)^{0.5}$. As is shown in equation (6.36), the column density of a density peak with a given δ_b (overdensity) is proportional to the following combination of parameters:

$$F \equiv \left(\frac{T_0}{10^4 K}\right)^{-0.7} \left(\frac{\Omega_b h_{100}^2}{0.0125}\right)^2 \left(\frac{J_{\rm HI}}{0.5}\right)^{-1} . \tag{6.41}$$

Hence, by equation (6.38), if F is rescaled by a certain factor (by changing T_0, Ω_b or $J_{\rm HI}$ or their combinations), the number of absorption lines is also changed by the same factor at an appropriately rescaled column density.

Suppose F is rescaled to F' such that $F' = rF$, then

$$\left.\frac{d^2 N'_{\rm Ly\alpha}}{dN'_{\rm HI} dz}\right|_{N'_{\rm HI}=rN_{\rm HI}} = \frac{1}{r} \left.\frac{d^2 N_{\rm Ly\alpha}}{dN_{\rm HI} dz}\right|_{N_{\rm HI}} . \tag{6.42}$$

It implies that if the column density distribution is a pure power law, then in a log-log plot of the number of absorption lines per unit column density per unit redshift versus column density, the straight line would simply be shifted to the right or left (or up/down) by rescaling. In reality, the column density distribution only approximately obeys a power law and so there should be a slight change of slope at any given column

Label	σ_8	n	$k_{\rm S}/Mpc^{-1}$	Ω_b	σ_0
CDM1	0.7	1.0	2.3	0.05	1.12
CDM2	0.7	1.0	2.3	0.06	1.12
CDM3	0.4	1.0	5.86	0.05	0.93
CDM4	0.7	0.7	3.58	0.05	0.93

Table 6-1: A list of all the CDM models discussed in this paper. All have $h_{100} = 0.5$. The spectral index of the power spectrum is n. Every model has $\Omega_b = 0.05$ except for CDM2, which has a higher baryon content and is shown in Fig. 6-6. The truncation scale $k_{\rm S}$ is defined by $k_{\rm S} = 1.5\, k_{\rm NL}$ (equation [6.20]). σ_0 is defined in equation (6.43). $\bar{z} = 3$ is assumed.

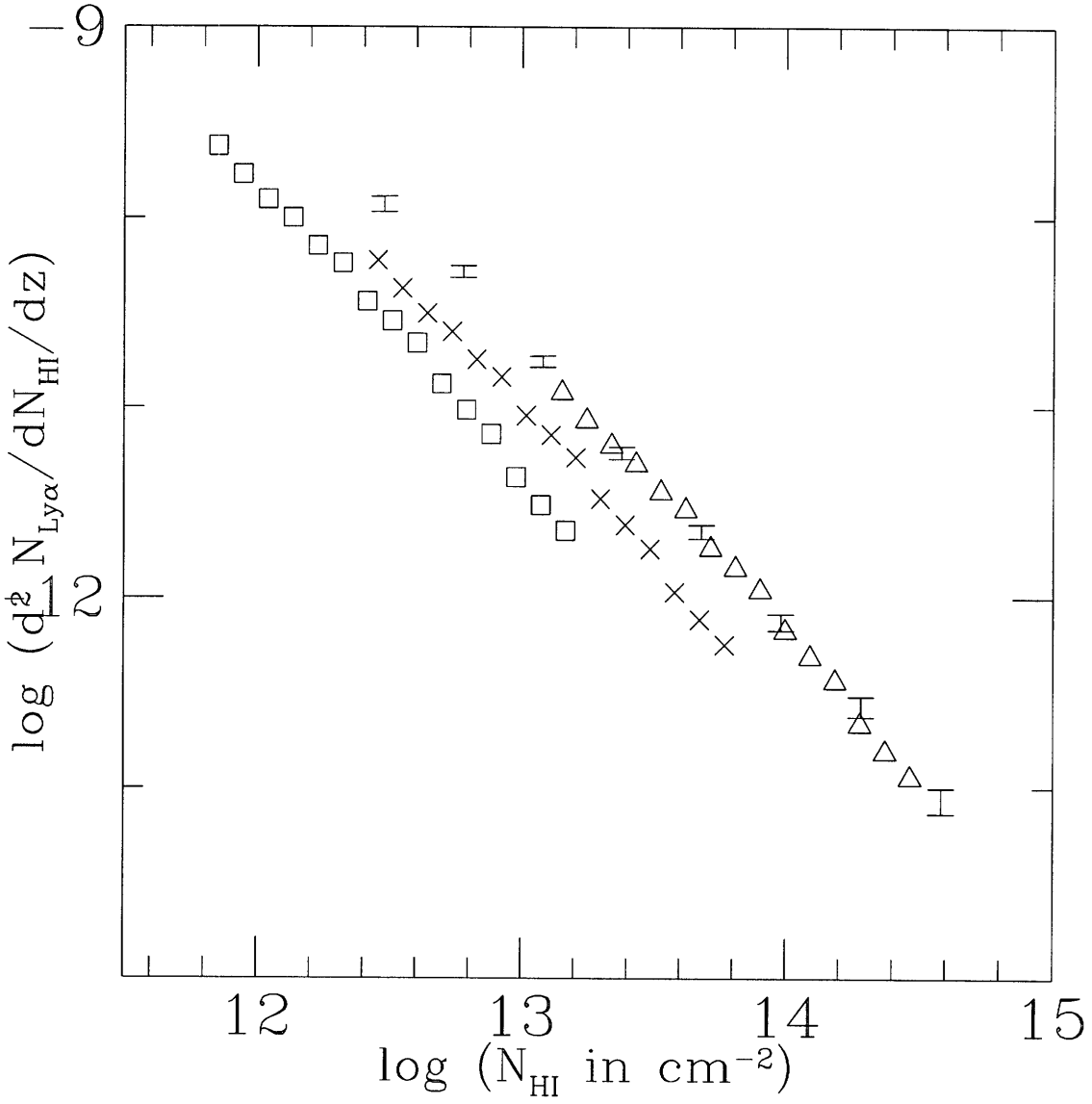

Fig. 6-9.— The column density distribution of the $\sigma_8 = 0.7$ CDM model with no tilt (CDM1 in Table 6-1), obtained using the Density-Peak-Ansätz and the truncated Zel'dovich approximation ($k_S = 2.3\,\mathrm{Mpc}^{-1}$). The redshift is $\bar{z} = 3$. Box size is 12.8 Mpc with grid spacing of 0.05 Mpc. Open squares: $F = 0.25$ (equation [6.41]). Crosses: $F = 1$. Open triangles: $F = 5$. All of them have the equation of state described by $\gamma = 1.5$ (eq. [6.28]). The points with error bars are the observational data at about $z = 3$ which have been corrected for incompleteness, taken from Hu et al. (1995).

density as a result of rescaling.

The effects of rescaling can be seen clearly in Fig. 6-9, where F is allowed to take the values 0.25, 1 and 5. Keeping $\Omega_b h_{100}^2 = 0.0125$ and $J_{\rm HI} = 0.5$, it corresponds to changing T_0 from about 72000 K to 1000 K. Alternatively, keeping T_0 and $\Omega_b h_{100}^2$ fixed at their canonical values (as shown in equation [6.41]), it corresponds to allowing $J_{\rm HI}$ to vary between 2 and 0.1. See Hui and Gnedin (1996) for a discussion of the dependence of T_0 on reionization history. T_0 is expected to fall within the range quoted above.

The conventional value of $\Omega_b h_{100}^2 = 0.0125$ has been challenged by recent measurements of light element abundance in high redshift absorption systems. Tytler and Burles (1996) obtain a value of 0.024, which for $T_0 = 10^4$ K and $J_{\rm HI} = 0.5$, implies $F = 3.686$, well within the range of F plotted in Fig. 6-9. The analysis of Hogan and Rugers (1996), on the other hand, favors the value 0.006, which means $F = 0.23$ for the same values of T and $J_{\rm HI}$. The lowest set of points in Fig. 6-9 has to be lowered further to accommodate this value of the baryon density.

The observational data are taken from Hu et al. (1995), measured at about redshift of about 3 and corrected for incompleteness. The error bar in the vertical direction is about the size of the solid square.

We note in passing that strictly speaking, altering Ω_b, in addition to rescaling the number of absorption lines as discussed above, also changes the transfer function in a non-trivial way. But the effect is very small for models in which the dark matter (non-baryons) dominate. In fact, using the BBKS (Bardeen et al. 1986) transfer function, which does not take into account the effect of baryons at all, instead of the Ma (1996) transfer function, which does take it into account for $\Omega_b h_{100}^2 = 0.0125$ with $h_{100} = 0.5$, gives essentially the same column density distribution for the range of column densities discussed here. For models where the baryon content is proportionally higher, Low-density Cold Dark Matter Models for instance, changing $\Omega_b h_{100}^2$ has a more pronounced effect on the transfer function.

6.6.2 Dependence on the Equation of State

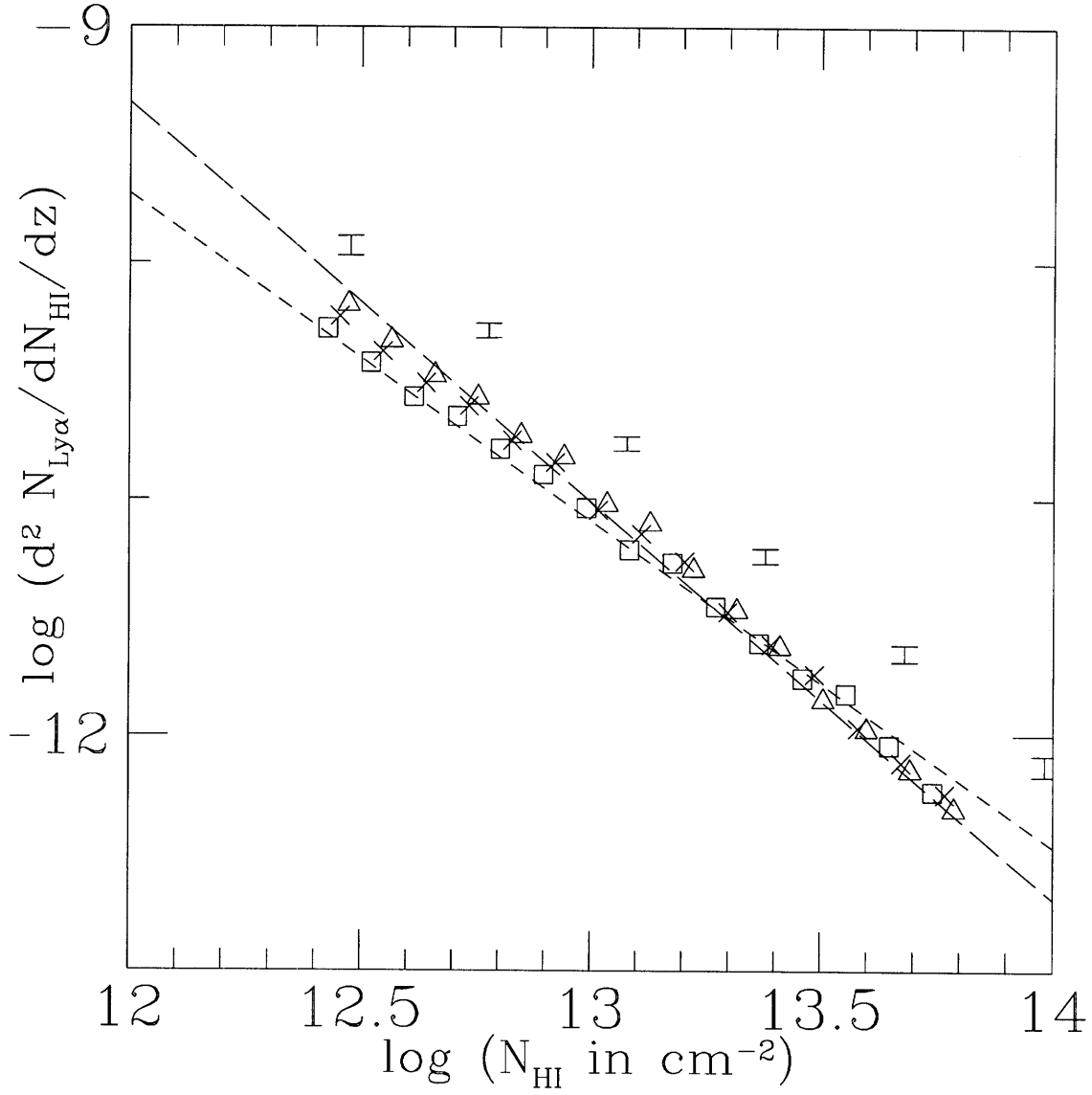

Fig. 6-10.— Column density distributions of the CDM1 model (Table 6-1) for three different equations of state (equation [6.28]). Open squares: $\gamma = 1.2$; crosses (same as crosses in Fig. 6-9): $\gamma = 1.5$; open triangles: $\gamma = 1.7$. $F = 1$ (defined in eq. [6.41]) for all three. Points with error-bars are the same observational data as in Fig. 6-9. Long-dashed and short-dashed lines give the approximate slopes (-1.69 and -1.38) as given in equation (6.47) for the open triangles and open squares respectively.

Let us hold fixed T_0, $\Omega_b h_{100}^2$ and $J_{\rm HI}$ at their canonical values as in equation (6.41) but allow the equation of state to change, for the same CDM model as above. As is pointed out in Sec. 6.3.2, the temperature-density relation for low enough overdensity is well-approximated by a power law where the power index is around 0.5, but can change slightly depending on the reionization history. We plot in Fig. 6-10 the column density distributions for $\gamma = 1.2, 1.5, 1.7$ where γ is defined by $T \propto (1+\delta_{\rm b})^{\gamma-1}$. It should adequately cover the possible range of γ (Hui & Gnedin 1996).

The first thing to notice is that the column density distribution remains almost the same for the three different values of γ. This is because γ affects column density through the power index of $(1+\delta_{\rm b})$, which is $2-0.7(\gamma-1)$ (equation [6.37]). The index does not change significantly for the range of γ considered. A larger index (smaller γ) means for a density peak with a given $1+\delta_{\rm b}$ (and its second derivative), the column density is larger or smaller depending on whether $1+\delta_{\rm b}$ is bigger or smaller than one. The net effect is to decrease the slope of the column density distribution. The effect, though very small for the values of γ plotted, can still be seen in Fig. 6-10. We also show the approximate slopes given by an analytical formula (equation 6.47) which will be discussed later. Note how the column density distribution does not exactly follow a power law but can be approximated by one.

Hence as a crude approximation, we conclude that the mean temperature, radiation intensity and baryon density mainly determine the overall normalization of the column density distribution. The equation of state, on the other hand, mostly affects the slope of the column density distribution but its effect is small for reasonable range of γ.

6.6.3 Dependence on σ_8

Putting $\gamma = 1.5$ and $F = 1$ (equation [6.41]), we now consider CDM models with different σ_8, in other words, different normalizations of the power spectrum. Plotted in Fig. 6-11 are two CDM models, one with $\sigma_8 = 0.7$, the other with $\sigma_8 = 0.4$, both without any tilt. Both have the same thermal and ionization parameters ($T_0 = 10^4$ K, $J_{\rm HI} = 0.5$ and $\gamma = 1.5$). The $\sigma_8 = 0.4$ model is not meant to be realistic but it serves

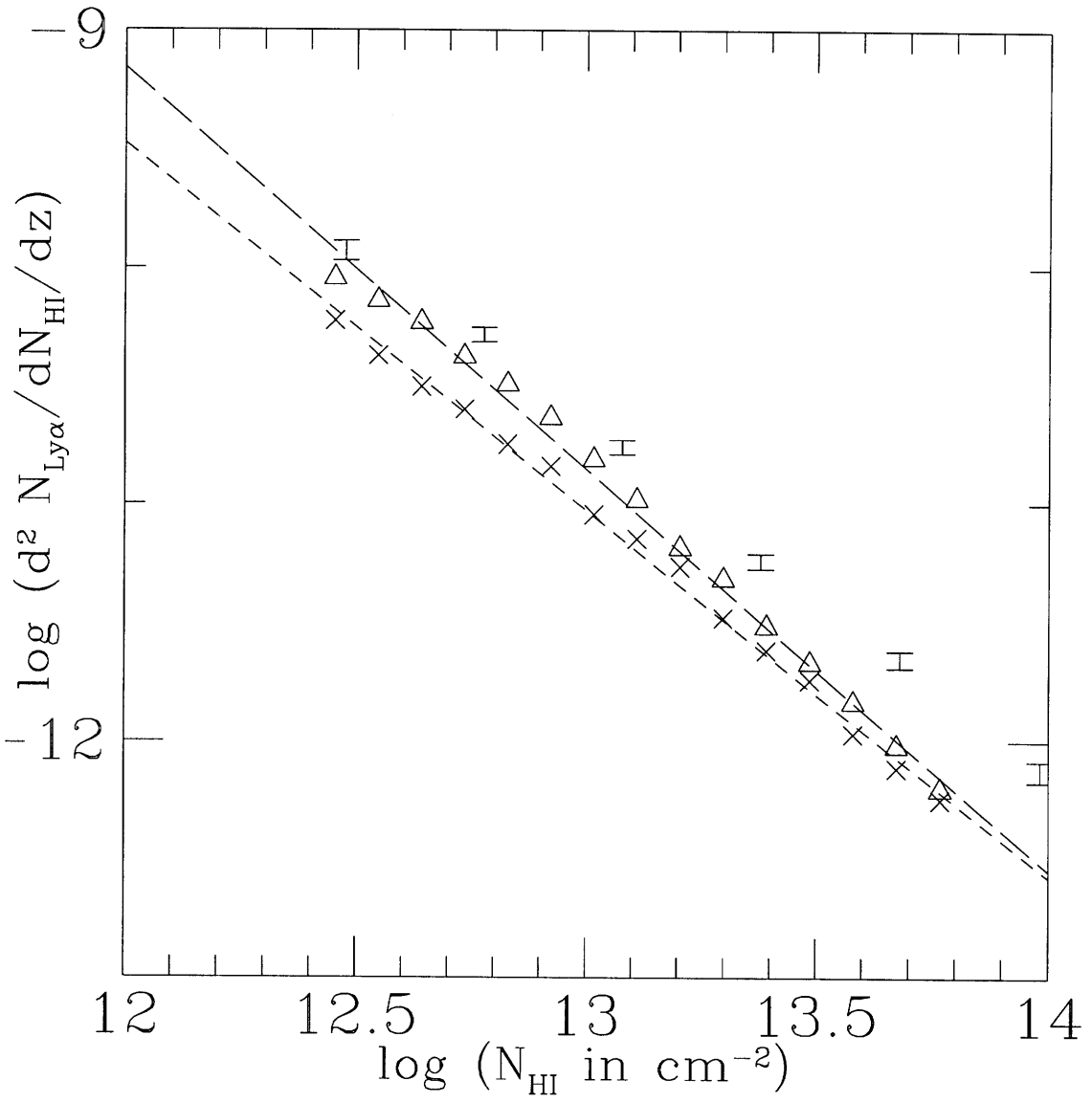

Fig. 6-11.— Solid squares are the same observational data as in Fig. 6-9. Open squares: the CDM1 model with $\sigma_8 = 0.7$ (see Table 6-1). Open triangles: the CDM3 model with $\sigma_8 = 0.4$. For both models, $F = 1$ (equation [6.41]) and $\gamma = 1.5$ (equation [6.28]) are adopted. The long-dashed and short-dashed lines give the approximate slopes (-1.69 and -1.55) as given in equation (6.47) for the open triangles and crosses respectively.

to illustrate the effect of choosing a different normalization.

It interesting to see that changing the normalization of the power spectrum alters the slope of the column density distribution. For column densities greater than about $10^{12.5}\,\mathrm{cm}^{-2}$, the lower σ_8 model has a steeper distribution. It indicates that the model is in a more "linear" state of evolution compared with the higher σ_8 model at a redshift of 3. There are proportionally more low density peaks compared to high density ones, hence the steeper distribution. For sufficiently low column densities, however, the absorption lines arise from very underdense regions which should be more common in the high σ_8 model. Hence at very low column densities, the high σ_8 model should win: it has more very low density peaks. Where this might occur we cannot tell from our simulations because of the limited resolution. For the range of column densities we can measure reliably, the slope of the column density distribution simply steepens as σ_8 is lowered.

6.6.4 Dependence on the Shape of the Power Spectrum

We investigate here the effect of changing the slope of the power spectrum while keeping σ_8 fixed. Fig. 6-12 shows two CDM models, both with $\sigma_8 = 0.7$, one without tilt and the other with the spectral index $n = 0.7$. Small scale power is suppressed in the $n = 0.7$ model. The effect on the column density distribution is very similar to lowering σ_8.

Another way to alter the shape of the power spectrum is to apply different amounts of smoothing to the *same* initial power spectrum.

In Fig. 6-13, we plot for the $\sigma_8 = 0.7$ CDM1 model (Table. 6-1) the results of three different smoothing scales. The crosses represent the result of using the smoothing scale according to the standard prescription (equation [6.21]). More smoothing (open triangles) has qualitatively the same effect as lowering σ_8 (Fig. 6-11). This should come as no surprise because smoothing means decreasing small scale power. Again, one ends up decreasing the number of high density peaks compared to low density peaks, hence resulting in a steeper column density distribution. Smoothing by $k_S = 1.15\,\mathrm{Mpc}^{-1}$ (open triangles) removes more structure than is necessary from

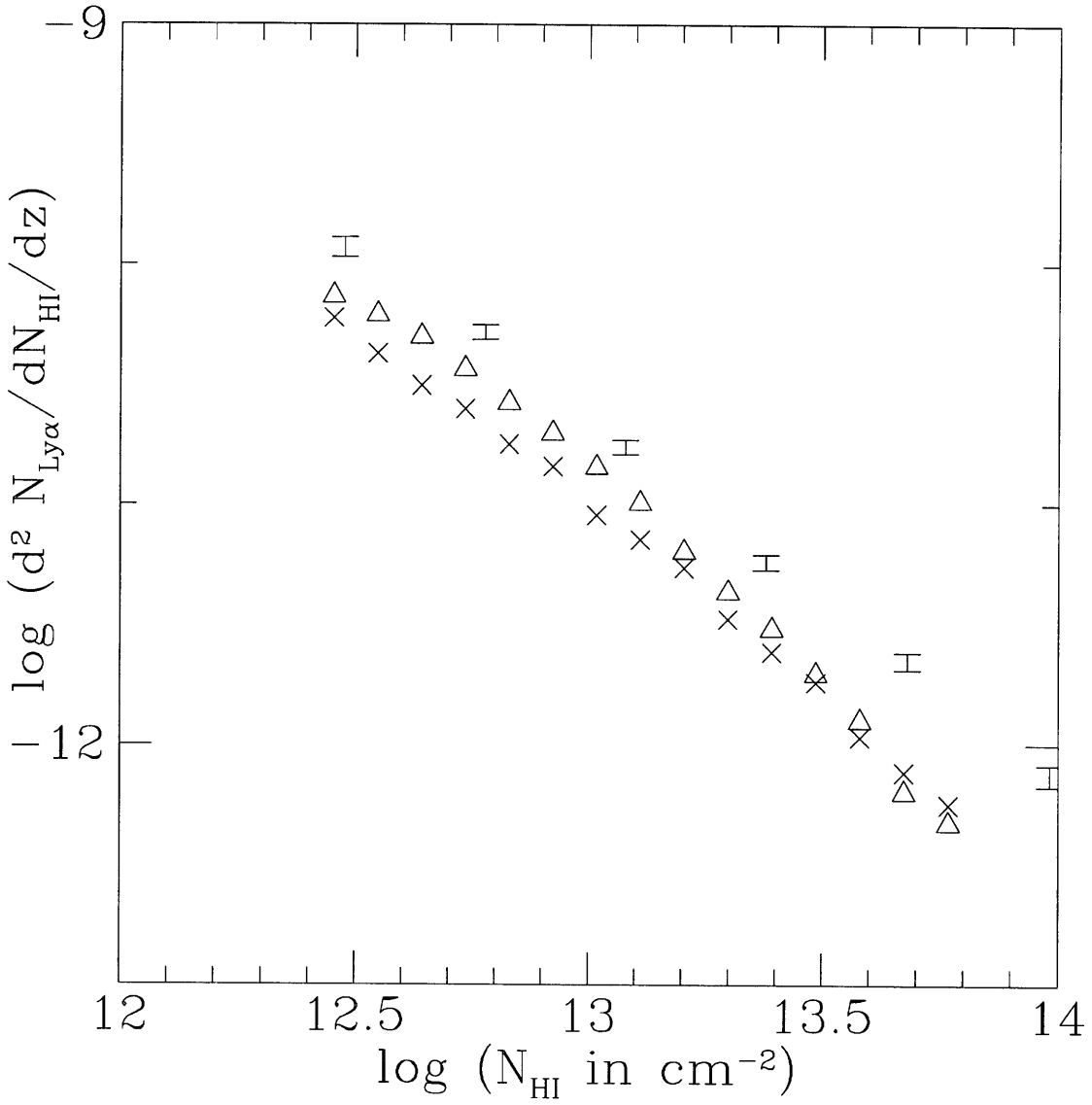

Fig. 6-12.— Points with error-bars are the same observational data as in Fig. 6-9. Crosses represent the $\sigma_8 = 0.7$ and no tilt CDM1 model as before (same as crosses in Fig. 6-9; see Table 6-1). Open triangles represent the column density distribution of the CDM4 model: $\sigma_8 = 0.7$ with tilt, the spectral index being 0.7. For both models, $F = 1$ (eq. [6.41]) and $\gamma = 1.5$ (equation [6.28]) are adopted.

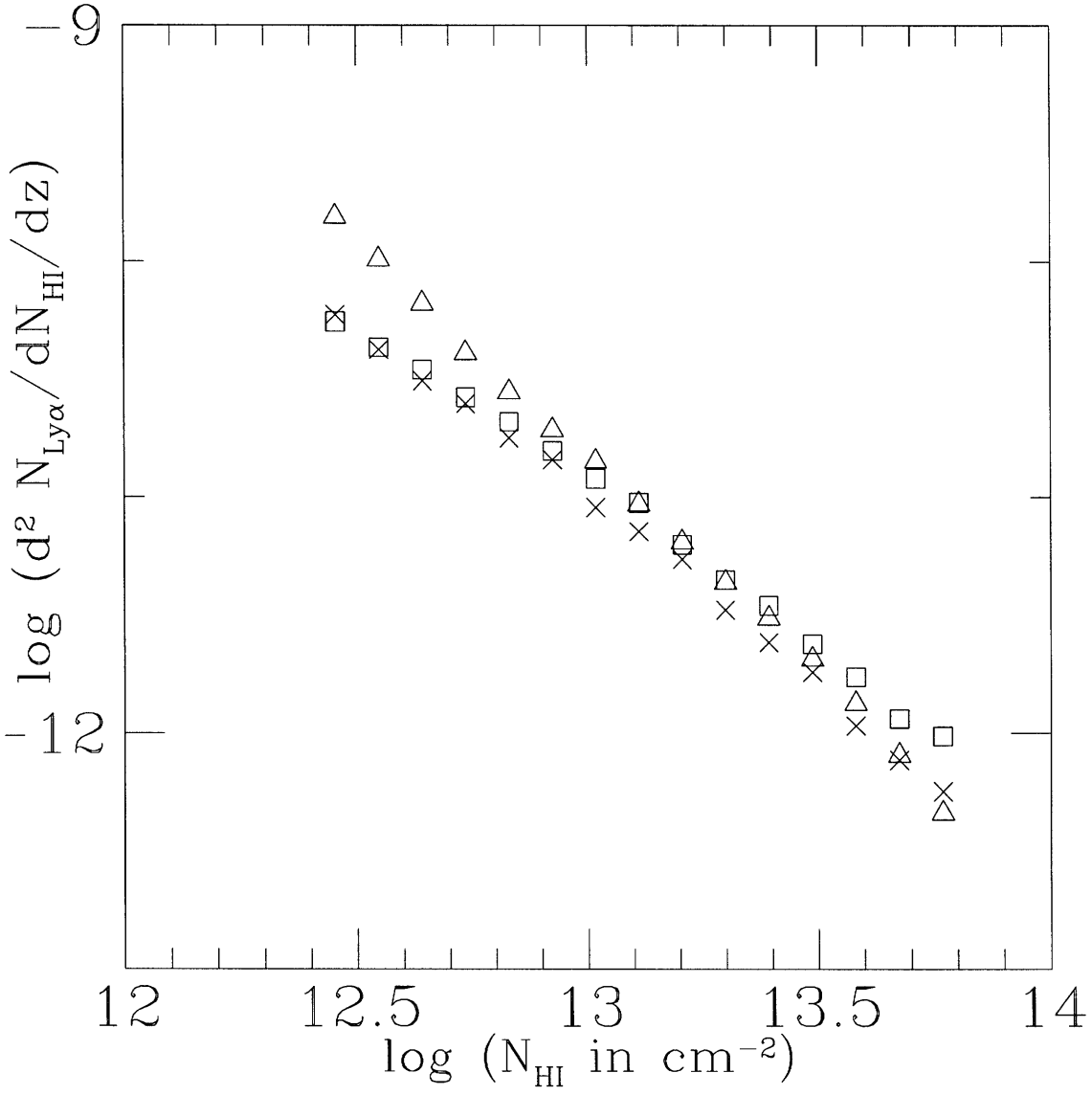

Fig. 6-13.— Column density distributions of the CDM1 model (Table 6-1) for three different initial smoothing scales. Crosses (same as crosses in Fig. 6-9): $k_S = 2.3$ Mpc^{-1}, which is the smoothing scale according to the standard prescription (equation [6.21]). Open triangles: $k_S = 1.15$ Mpc^{-1}. Open squares: $k_S = 8.4$ Mpc^{-1}, which is the Jeans scale for $T_0 = 1$ and $\gamma = 1.5$. We adopt $F = 1$ (eq. [6.41]) and $\gamma = 1.5$ (eq. [6.28]) for all three cases.

the point of view of optimizing the truncated Zel'dovich approximation. Smoothing by $k_S = 8.4\,\mathrm{Mpc}^{-1}$ has the opposite effect of making the column density distribution flatter, but it should be kept in mind that severe orbit-crossing occurs because not enough smoothing is performed in this case. The comparison with a full hydrodynamic simulation in Fig. 6-6 lends support to the standard prescription for choosing k_S (equation [6.21]).

To understand the overall effect of changing the power spectrum (both its shape and normalization), we plot in Fig. 6-14 the smoothing scale k_S versus the following quantity:

$$\sigma_0 = D_+(t)\sqrt{\int_0^\infty 4\pi k^2 P(k) e^{-(k/k_S)^2} dk}\,. \tag{6.43}$$

which is the linear rms fluctuation of a density field smoothed with a spherically symmetric Gaussian window with radius $1/k_S$. The linear growth factor $D_+(t)$ is equal to $(1+\bar{z})^{-1}$ if $D_+ = 1$ today is assumed. It is one way to represent the amount of fluctuation around comoving scale $1/k_S$. Three power spectra are shown: CDM1, CDM3 and CDM4 (see Table 6-1), all at $\bar{z} = 3$. The Jeans scale is around $8\,\mathrm{Mpc}^{-1}$ (Sec.6.3.1), below which we expect the intergalactic medium to be smooth (except for very small scales which have collapsed to form stars, etc) and so does not contribute significantly to absorption lines. For scales large enough such that σ_0 is smaller than, say 0.1, the density fluctuations are essentially linear. These are scales that would collapse later to form the larger scale structure we see today. The average amplitude of fluctuations associated with them is so small at $z = 3$ that the intergalactic medium is also smooth on those scales. This leaves around one to two decades of length scales below $8\,\mathrm{Mpc}^{-1}$, on which the density fluctuations contribute to the absorption lines of column densities of interest in this paper. Models with similar σ_0 on these scales are therefore expected to have similar column density distributions. (Gnedin and Hui 1996 reached similar conclusions using the lognormal approximation). From Fig. 6-14, one can then understand why the $\sigma_8 = 0.4$ model and the $n = 0.7$, $\sigma_8 = 0.7$ have qualitatively similar column density distribution. Namely, both have smaller σ_0 on

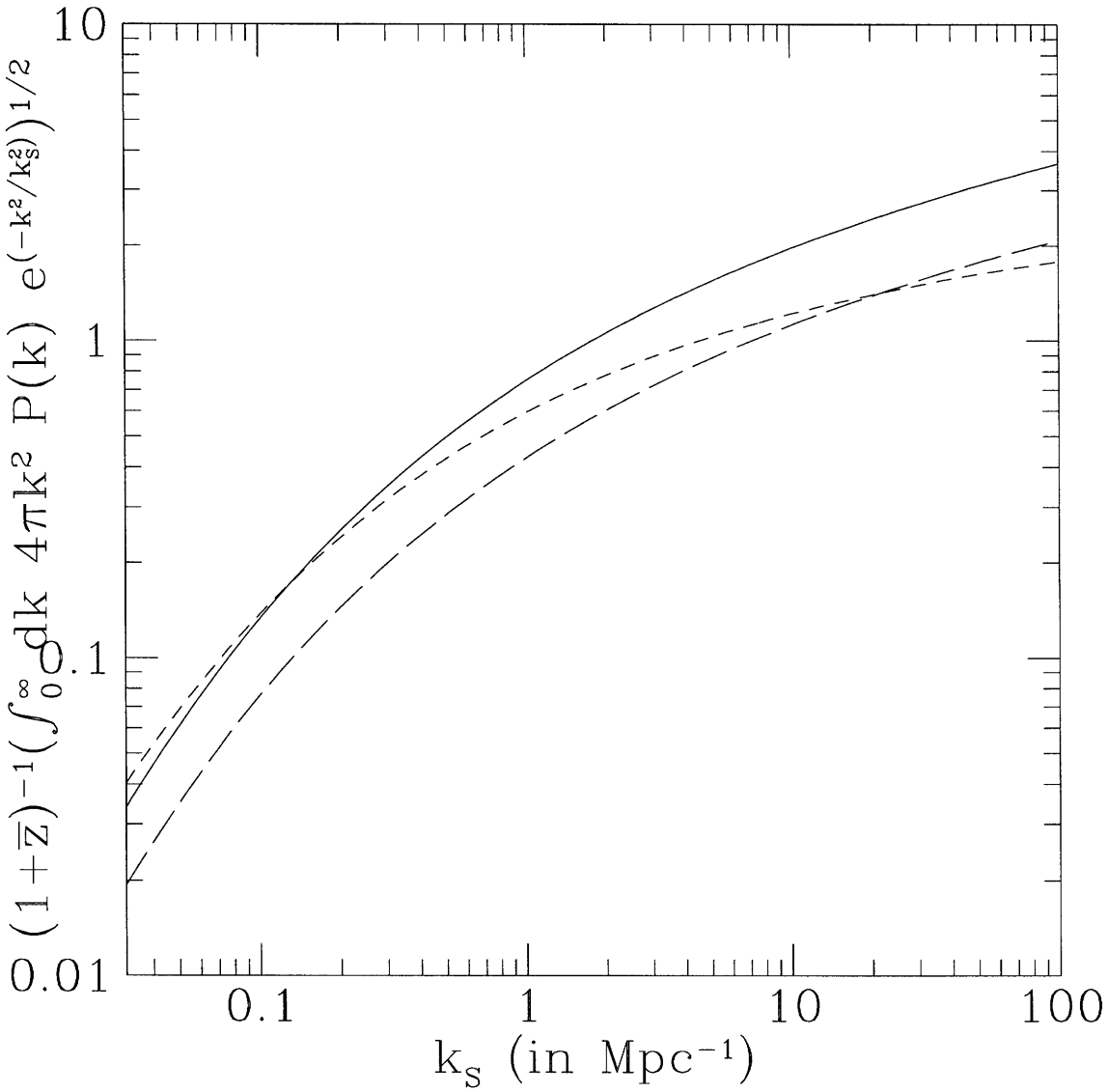

Fig. 6-14.— σ_0 versus k_S (equation [6.43]) at $\bar{z} = 3$. Solid line: $\sigma_8 = 0.7$ CDM with no tilt (CDM1); short dash: $\sigma_8 = 0.7$ CDM with tilt, spectral index being 0.7 (CDM3); long dash: $\sigma_8 = 0.4$ CDM with no tilt (CDM4).

the relevant scales compared with the $\sigma_8 = 0.7$ no-tilt-model and hence both have steeper column density distribution for column densities higher than $10^{12.5}$ cm^{-2}.

6.7 The Slope of the Column Density Distribution

It has been shown that while the normalization of the column density distribution is influenced by the thermal and ionization states of the intergalactic medium which are not well-constrained, the slope is sensitive to the amount of small scale power and depends weakly on the effective equation of state. We develop a quantitative expression for the later relation here.

From equations (6.36) and (6.37), it can be seen that the column density $N_{\rm HI}$ is proportional to $(1+\delta_b)^{2-0.7(\gamma-1)}$ times $1/\sqrt{\xi''}$, which basically defines a length scale. Taking into account the correlation between this length scale and the overdensity, we find from our simulations (which use the Zel'dovich approximation) a useful approximate relation for column densities between about $10^{12.5}$ and $10^{14.5}$ cm^{-2}:

$$N_{\rm HI} \propto (1+\delta_b)^{1.05-0.7(\gamma-1)}, \tag{6.44}$$

which means the length scale $1/\sqrt{\xi''}$ is approximately proportional [6] to $(1+\delta_b)^{-0.95}$.

Now, since we are interested in the slope of the column density distribution, the relevant quantity to consider is:

$$-\beta \equiv \frac{d\ln \frac{d^2N_{\rm Ly\alpha}}{dN_{\rm HI}dz}}{d\ln N_{\rm HI}} = -1 + \frac{m}{1.05-0.7(\gamma-1)}\ ,\ m \equiv \frac{d\ln \int_{-\infty}^{0} d\xi''|\xi''|P(\xi,\xi'=0,\xi'')}{d\xi}. \tag{6.45}$$

The equality follows from equations (6.44), (6.36), (6.38) and (6.40) and noting that $\xi \equiv 1+\delta_b$. The column density distribution can be approximated by $N_{\rm HI}^{-\beta}$ if β defined above is only weakly dependent on $N_{\rm HI}$.

[6] A log-log plot of $1/\sqrt{\xi''}$ versus $(1+\delta_b)$ actually shows a lot of scatter but equation (6.44) appears to capture the overall dependence of $N_{\rm HI}$ on $1+\delta_b$.

Now, lacking an analytical expression for P under the Zel'dovich approximation, we can nonetheless guess what the general properties of the derivative $d\ln\int_{-\infty}^{0}d\xi''|\xi''|P/d\xi$ are. First of all, the derivative depends on ξ because the integral $\int_{-\infty}^{0}d\xi''|\xi''|P$ cannot be a simple power law in ξ. This is because we expect the integral (and P itself) to vanish for very large and very small ξ and peak at some intermediate ξ. Suppose $\xi_{\rm pk}$ is the value of ξ where the integral $\int_{-\infty}^{0}d\xi''|\xi''|P$ reaches its maximum value, then the column densities we are interested in must correspond to $\xi > \xi_{\rm pk}$ where $d\ln\int_{-\infty}^{0}d\xi''|\xi''|P/d\xi$ is negative. This is based on the knowledge that the computed (as well as observed) slope in equation (6.45) is less than -1 (the factor $1.05-0.7(\gamma-1)$ is positive). Furthermore, as we have noted before, we expect the relative number of high density to intermediate density (for ξ larger than ξ_{rmpk}) peaks to be lower for models with less power, which means $d\ln\int_{-\infty}^{0}d\xi''|\xi''|P(\xi,\xi'=0,\xi'')/d\xi$ is more negative. Lastly, since the Zel'dovich displacement field is Gaussian in Lagrangian space, we expect on quite general grounds that P depends on the power spectrum through the 3 parameters: σ_0 as defined in equation (6.43) and σ_1 and σ_2 as defined as follows (see Bardeen et al. 1986):

$$\sigma_1 = D_+(t)\sqrt{\int_0^\infty 4\pi k^4 P(k)e^{-(k/k_{\rm S})^2}dk}\,, \quad \text{and } \sigma_2 = D_+(t)\sqrt{\int_0^\infty 4\pi k^6 P(k)e^{-(k/k_{\rm S})^2}dk}\,, \tag{6.46}$$

where D_+ is the linear growth factor which is equal to $(1+\bar{z})^{-1}$ for a universe at critical matter density.

In the last section, we have tested two CDM models that have the same shape of the power spectrum but different normalizations (see Fig. 6-11). These models have the same values of σ_1/σ_0 and σ_2/σ_0, but different σ_0. One can then try to find the σ_0 dependence of $d\ln\int_{-\infty}^{0}d\xi''|\xi''|P(\xi,\xi'=0,\xi'')/d\xi$ for these models (σ_1/σ_0 and σ_2/σ_0 being fixed). We find the following form fits the slope of the computed column density distribution reasonably well (Fig. 6-11):

$$\beta = 1 + \frac{1}{1.05-0.7(\gamma-1)}(1-0.55\sigma_0)\,. \tag{6.47}$$

where β satisfies equation (6.45). An example of how well it describes the variation of the slope with a change in the equation of state can be found in Fig. 6-10.

The above form for β displays the expected dependence of the slope on σ_0, namely steeper slope with lower σ_0. However, we don't expect it to hold in general because the dependence on σ_1 and σ_2 is ignored. In fact, it fails to describe accurately the slope of the column density distributions of the CHDM models we are going to study.[7] Nonetheless, equation (6.45) should give a good description of how the slope of the column density distribution varies with γ. We can also see that it is the combination of γ around 1.5 and m around 0.35 in equation (6.45) that gives $\beta \sim 1.5$. It should be kept in mind, however, that the column density distribution is not expected to obey an exact power law, based on the general considerations stated before.

6.8 The Column Density Distribution for CHDM Models

In Sec. 6.6, we learn that while T_0, $J_{\rm HI}$ and $\Omega_b h_{100}^2$ play a major role in determining the normalization of the column density distribution, the amount of power on comoving scales between around 1 $\rm Mpc^{-1}$ and 10 $\rm Mpc^{-1}$ is primarily responsible for its shape or slope. The equation of state also has a small effect on the slope of the column density distribution.

We apply these insights to study a group of CHDM models. They are all $\Omega_0 = 1$ models with $\Omega_b = 0.05$. Both the $\Omega_\nu = 0.2$ and $\Omega_\nu = 0.1$ versions are considered. They have been shown to give good agreement with the observational data on large scales (k around $0.02 - 0.4$ $\rm Mpc^{-1}$) (see Fig. 6 and 7 of Ma 1996). The $\Omega_\nu = 0.3$ models seem to conflict with observed abundance of damped Lyman-α systems, which correspond to roughly k around $0.1 - 1.0\,\rm Mpc^{-1}$ comoving in the linear power spectrum. (Mo & Miralda-Escudé 1994; Kauffmann & Charlot 1994;

[7] An alternative analytic expression for β is being developed. See Hui and Gnedin 1996 for further discussion / improved treatment.

Label	Ω_ν	h_{100}	n	$Q_{\rm rms}/\,\mu$K	T/S	$k_{\rm S}/\,{\rm Mpc}^{-1}$
A1	0.1	0.5	0.95	18.5	$7(1-n)$	3.8
A2	0.2	0.5	0.95	18.5	$7(1-n)$	8.4
A3	0.3	0.5	0.95	18.5	$7(1-n)$	not apply
B1	0.1	0.5	0.9	19.2	0	3.3
B2	0.2	0.5	0.9	19.2	0	8.4
C1	0.1	0.65	0.9	19.2	$7(1-n)$	3.1
C2	0.2	0.65	0.9	19.2	$7(1-n)$	10.9
D1	0.1	0.65	0.8	20.5	0	2.8
D2	0.2	0.65	0.8	20.5	0	10.9

Table 6-2: A list of all the CHDM models discussed in this paper. All have $\Omega_b = 0.05$. The parameters are defined as follows: Ω_ν is the density parameter in neutrino, n is the spectral index of the power spectrum, Q_rms is the COBE quadrupole in μK and T/S is the tensor to scalar ratio. The smoothing wavenumber $k_{\rm S}$ for each $\Omega_\nu = 0.1$ model is $1.5\,k_{\rm NL}$ (equation [6.20]) and $k_{\rm S}$ for each $\Omega_\nu = 0.2$ model is the Jeans scale for $\gamma = 1.5$, $T_0 = 10^4\,K$ and the corresponding h_{100} (see Sec. 6.3.2). No simulation is run for $A3$, so no $k_{\rm S}$ is listed. All parameters and power spectra are taken from Ma (1996).

Ma & Bertschinger 1994). We list in Table 6-2 the CHDM models considered in this paper. We include one $\Omega_\nu = 0.3$ CHDM model for the sake of comparison. As is shown convincingly by Ma (1996), all models need some amount of tilt to match observations.

We compute as before the column density distribution for each model using the Density-Peak-Ansätz and the Zel'dovich approximation with appropriate smoothing. The power spectrum for each CHDM model is taken from Ma (1996).

As we have discussed before in the context of the CDM models, a plot of rms smoothed linear density fluctuation σ_0 versus smoothing scale $k_{\rm S}$ (equation [6.43]) is a very good indicator of what column density distribution to expect. This is done in Fig. 6-15 for the CHDM models tabulated in Table 6-2. The no-tilt $\sigma_8 = 0.7$ CDM model is also plotted for comparison.

Because of neutrino free streaming, all CHDM models have less power than the CDM model on small scales. Those with more neutrino content ($\Omega_\nu = 0.2$) have even less power than the others. In fact, the $\Omega_\nu = 0.2$ models have $\sigma_0 < 1$ on all scales

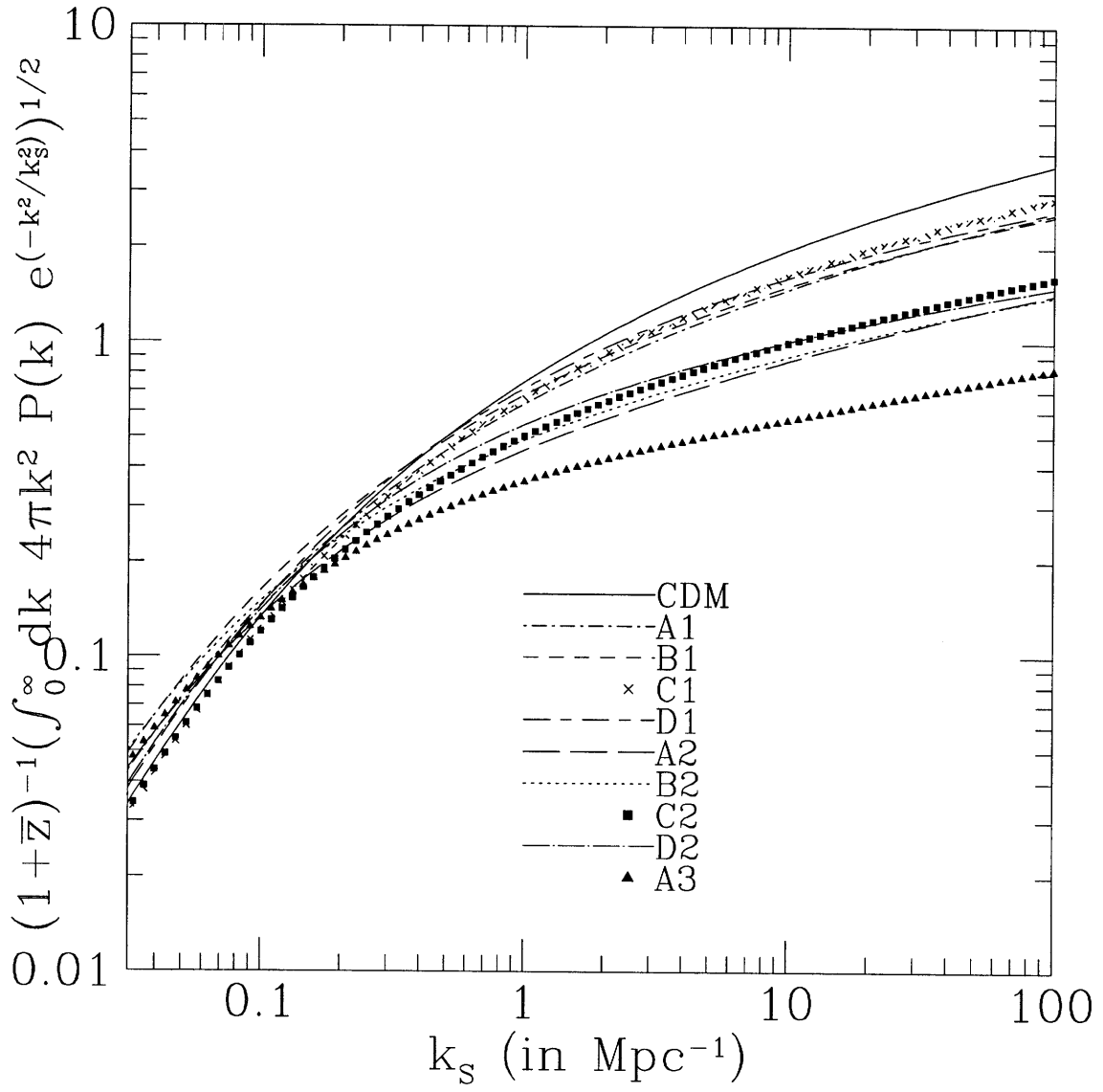

Fig. 6-15.— Going from the top, the solid line is the same $\sigma_8 = 0.7$, $h_{100} = 0.5$ CDM model with no tilt as in Fig. 6-9; the next four sets of points/lines close together are all $\Omega_\nu = 0.1$ CHDM models, A1, B1, C1 and D1 in Table 6-2 ; the next four sets are all $\Omega_\nu = 0.2$ CHDM models, A2, B2, C2 and D2; the last set, open squares, is an $\Omega_\nu = 0.3$ CHDM model, A3. All are shown at $\bar{z} = 3$.

larger than the Jeans scale ($k_S < k_J$). One expects the Zel'dovich approximation to work particularly well for these models because the amount of orbit-crossing will not be significant, even without initial truncation.

This is borne out by the next test: we compute the column density distribution for one CHDM model B2 and examine the effect of choosing different smoothing scales. The result is plotted in Fig. 6-16. The column density distribution in the range plotted does not change much at all for the three different smoothing scales plotted. Contrast this with the case of $\sigma_8 = 0.7$ CDM (Fig. 6-13) where the column density distribution is more sensitive to changes in the smoothing scale. That's why the truncation scale has to be chosen with some care: not too small (k_S too big) so that too much orbit crossing has occurred and not too large so that too much small scale structure is erased. We have shown the standard prescription (Sec. 6.3.1) to be a good one in Sec. 6.5.

For the CHDM model considered (in fact, it holds true for all other $\Omega_\nu = 0.2$ models here), the amount of small scale power is so insignificant that excluding them by smoothing does not affect the overall column density distribution at all (except possibly that one loses the small scale fluctuations that can give rise to very low column density absorption i.e. lower than our resolution limit). We have also done similar tests for the $\Omega_\nu = 0.1$ models, their response to changes in the truncation scale is somewhere between the $\sigma_8 = 0.7$ CDM model and the $\Omega_\nu = 0.2$ CHDM models, as can be expected based on their difference in Fig. 6-15.

We adopt the following truncation scales for the CHDM models. For the $\Omega_\nu = 0.1$ models, the standard prescription described in Sec. 6.3.1 is used (i.e. $k_S = 1.5\,k_{NL}$). The $\Omega_\nu = 0.2$ models, according to the above prescription, would have truncation scales less than the Jeans length ($k_S > k_J$) and so by the arguments presented in Sec. 6.3.1, $k_S = k_J$ is adopted. Again, we emphasize that for this class of models that have relatively little power on small scales, the precise truncation scale is not important. A summary of the truncation scales for all models can be found in Table 6-2.

The CHDM models with $\Omega_\nu = 0.1$ are plotted in Fig. 6-17. Values of F that

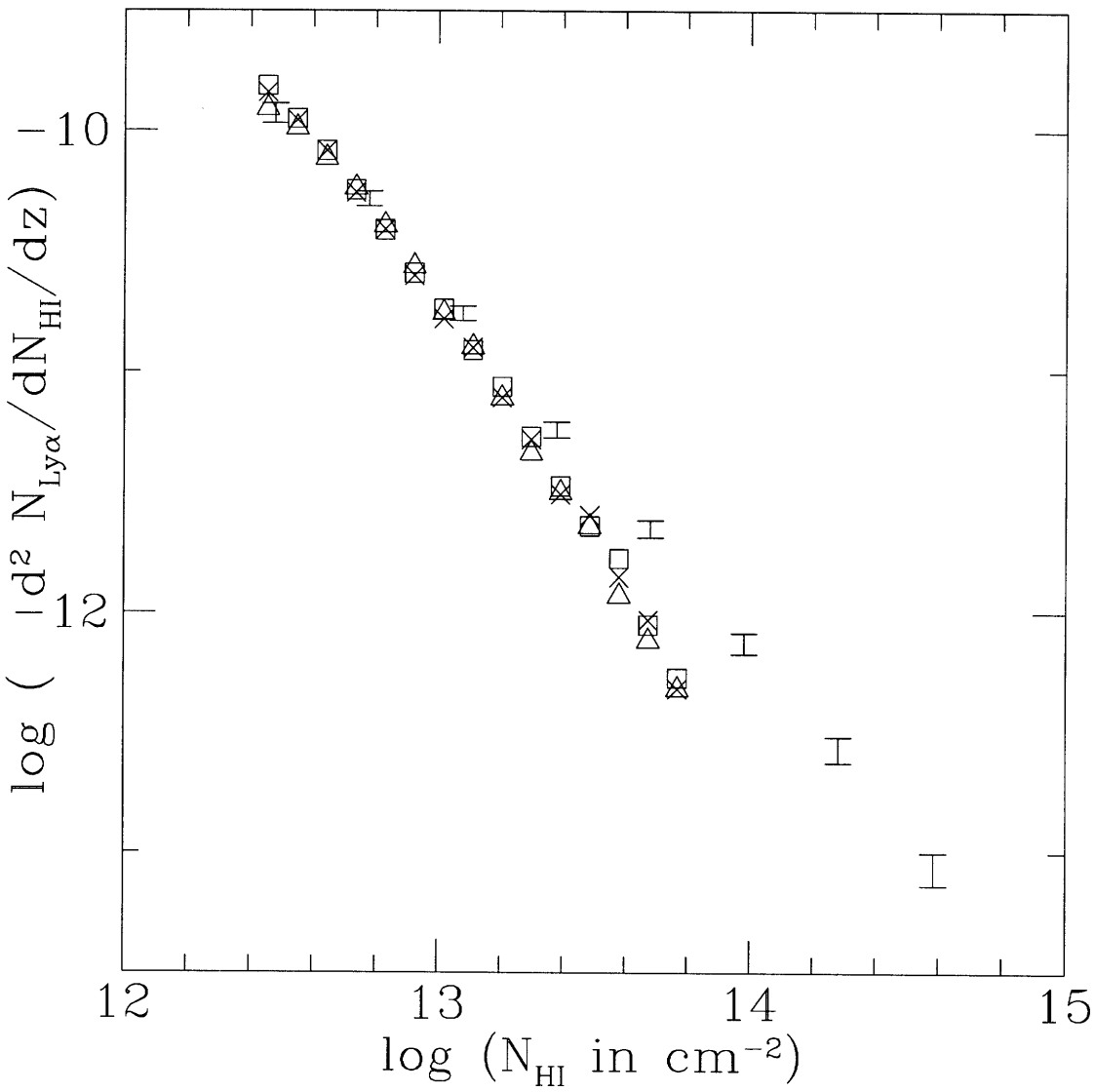

Fig. 6-16.— Column density distributions of the B2 CHDM model (see Table 6-2) for three initial smoothing scales. Crosses: $k_S = 19.2\,\mathrm{Mpc}^{-1}$ (standard truncation prescription, $k_S = 1.5\,k_{NL}$ according to eq. [6.20]). Open triangles: $k_S = 8.4\,\mathrm{Mpc}^{-1}$ (Jeans scale for $T_0 = 10^4$ and $\gamma = 1.5$). Open squares: no smoothing at all. Points with error-bars are the observational data as in Fig. 6-9. For all models, $F = 1$ (equation [6.41]) and $\gamma = 1.5$ (equation [6.28]) are used.

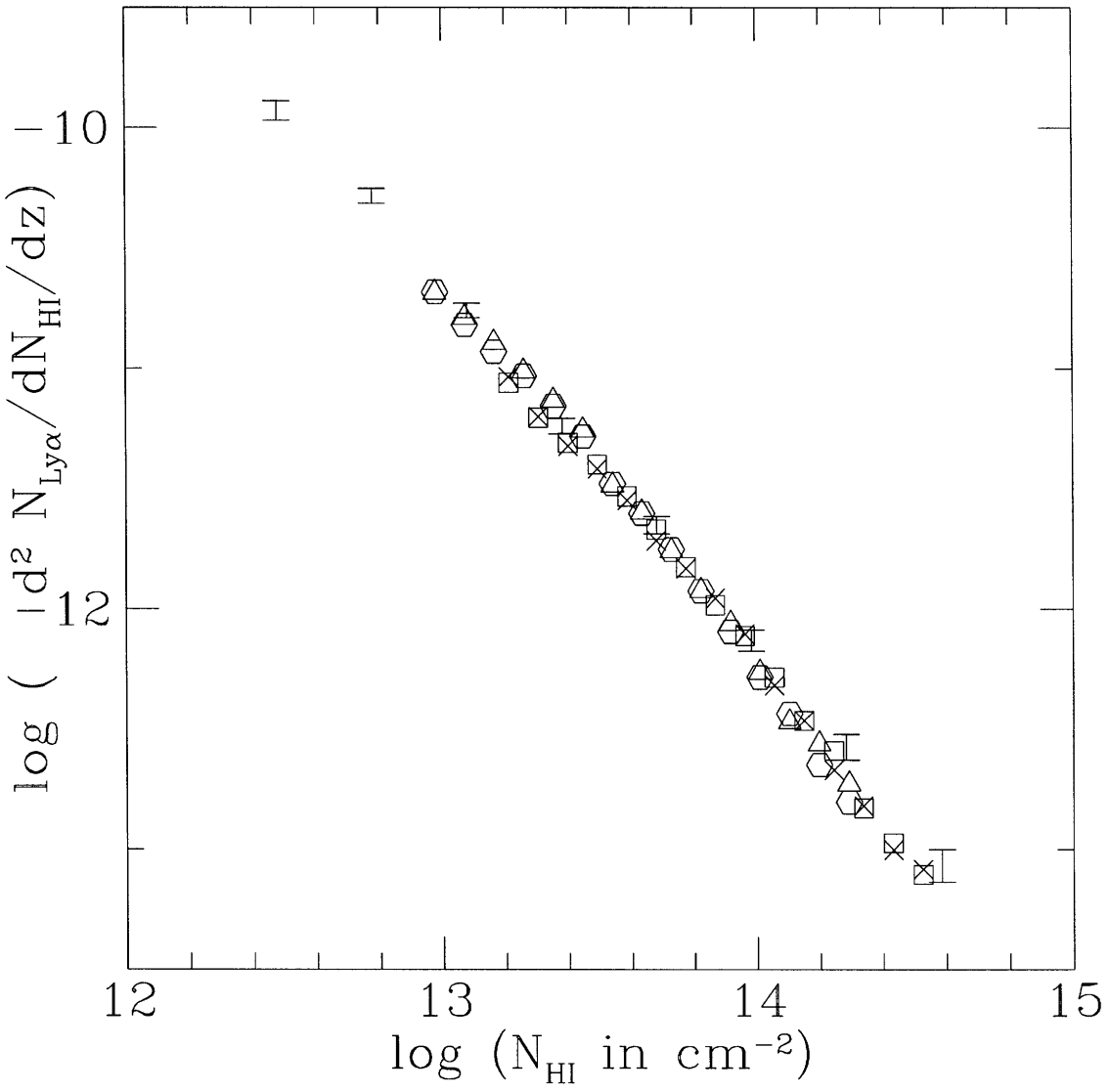

Fig. 6-17.— Column density distributions for four $\Omega_\nu = 0.1$ CHDM models. Points with error-bars are the observational data as in Fig. 6-9. All models have $\Omega_\nu = 0.1$. We use $\gamma = 1.5$ in the equation of state for all of them (equation [6.28]). Table 6-2 contains descriptions of each of the following model. Open hexagons: B1, $F = 3.33$. Open triangles: A1, $F = 3.33$. Crosses: D1, $F = 5.7$. Open squares: C1, $F = 5.7$. F is defined in equation (6.41).

give reasonable match to the observational data are chosen for each model. Note how the low-Hubble-constant-models ($h_{100} = 0.5$) requires a slightly lower F (equation [6.41]) than the higher-Hubble-constant-models. The equation of state is chosen to be the same for all models ($\gamma = 1.5$, see equation [6.28]). The level of agreement with the observational data, for the given choices of parameters, is satisfactory. Notice how the low Hubble-constant ($h_{100} = 0.5$)models tend to have steeper column density distributions, because they have less power on the relevant scales (see Fig. 6-15). Their slopes can be brought into better agreement with that of the observational data if a smaller γ is used.

For the $\Omega_\nu = 0.2$ models, we cannot find values of F that gives the same level of agreement with observations for $\gamma = 1.5$. Two examples are shown in Fig. 6-19 and Fig. 6-18. Both have $h_{100} = 0.5$ and small amounts of tilt. For each, three sets of theoretical predictions are plotted, one for each value of F, 1, 2.5 or 5. For $\Omega_b h_{100}^2 = 0.0125$, the conventional Big-bang nucleosynthesis value, and $T_0 = 10^4 K$, they correspond to radiation intensity $J_{\rm HI}$ of 0.5, 0.2 and 0.1 (equation [6.41]). As we have shown before, changing F mainly shifts the sets of points without altering the slope significantly. For the column density between about $10^{12.5}$ and $10^{14.5}\,{\rm cm}^{-2}$, the slope of the predicted distribution seems to be too steep compared to the observational data.

Two other $\Omega_\nu = 0.2$ CHDM models, which share a higher Hubble constant ($h = 0.65$), are shown in Fig. 6-20 and Fig. 6-21. The slope of their column density distribution is not as steep as the previous ones. This is expected because the higher Hubble constant models have slightly more power on relevant scales, as is evident in Fig. 6-15. In fact, one might argue that the middle set of points in each figure, the one having $F = 3.57$, match the observational data reasonably well if both observational and theoretically errors are taken into account. However, it is still true these two models predict a steeper column density distribution for $N_{\rm HI}$ between about $10^{12.5}$ and $10^{14.5} cm^{-2}$, compared to the $\Omega_\nu = 0.1$ CHDM models (Fig. 6-17).

It is not hard to understand the column density distributions of the CHDM models presented if one goes back to Fig. 6-15. The $\Omega_\nu = 0.2$ models have less power

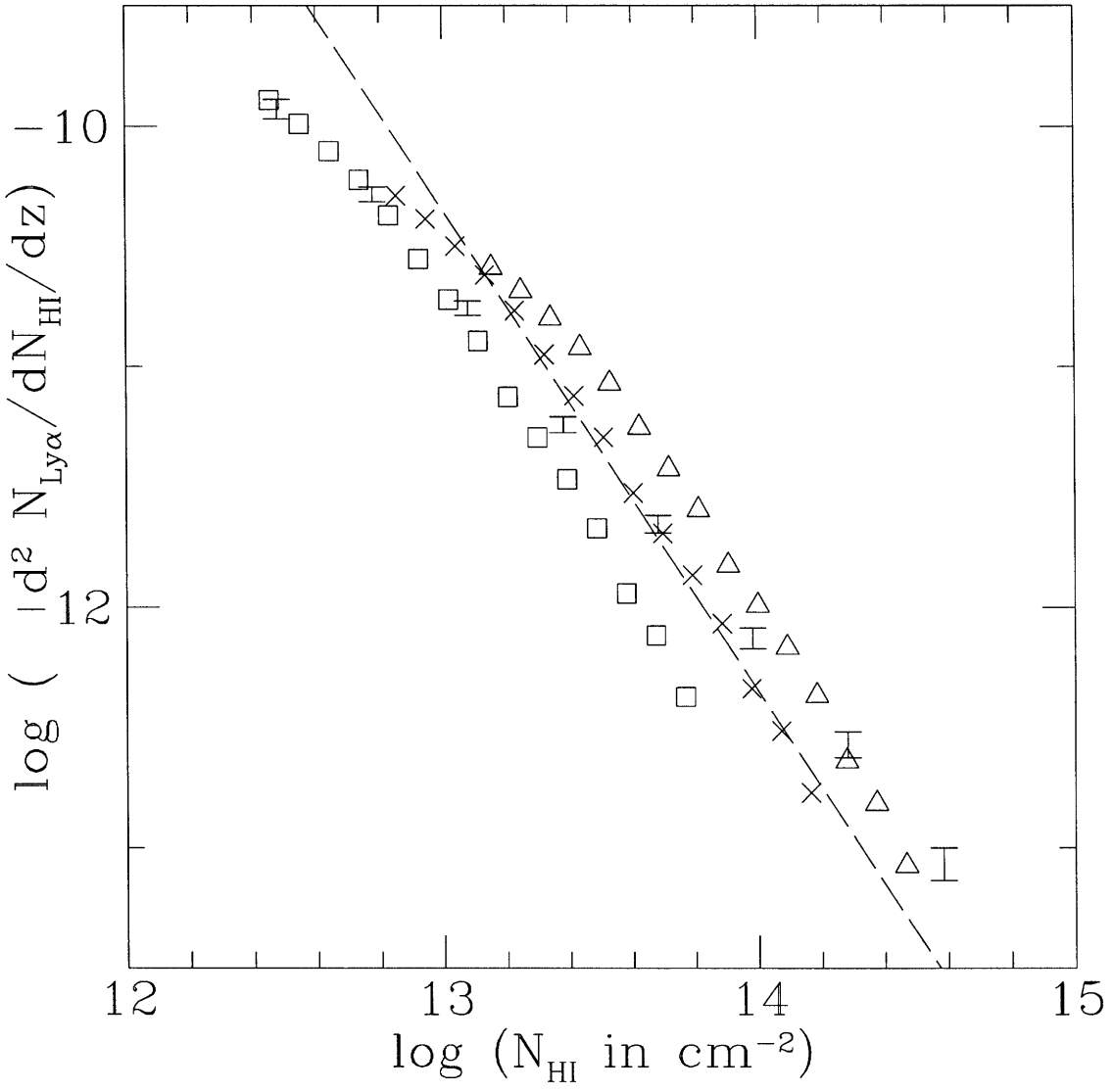

Fig. 6-18.— The column density distribution for the $A2$ CHDM model (Table 6-2). Three values of F (equation [6.41]) are shown: $F = 1$ (open squares), $F = 2.5$ (crosses) and $F = 5$ (open triangles). We choose $\gamma = 1.5$ for all three (equation [6.28]). Points with error-bars are the observational data as in Fig. 6-9. The dashed line has a slope of -2.

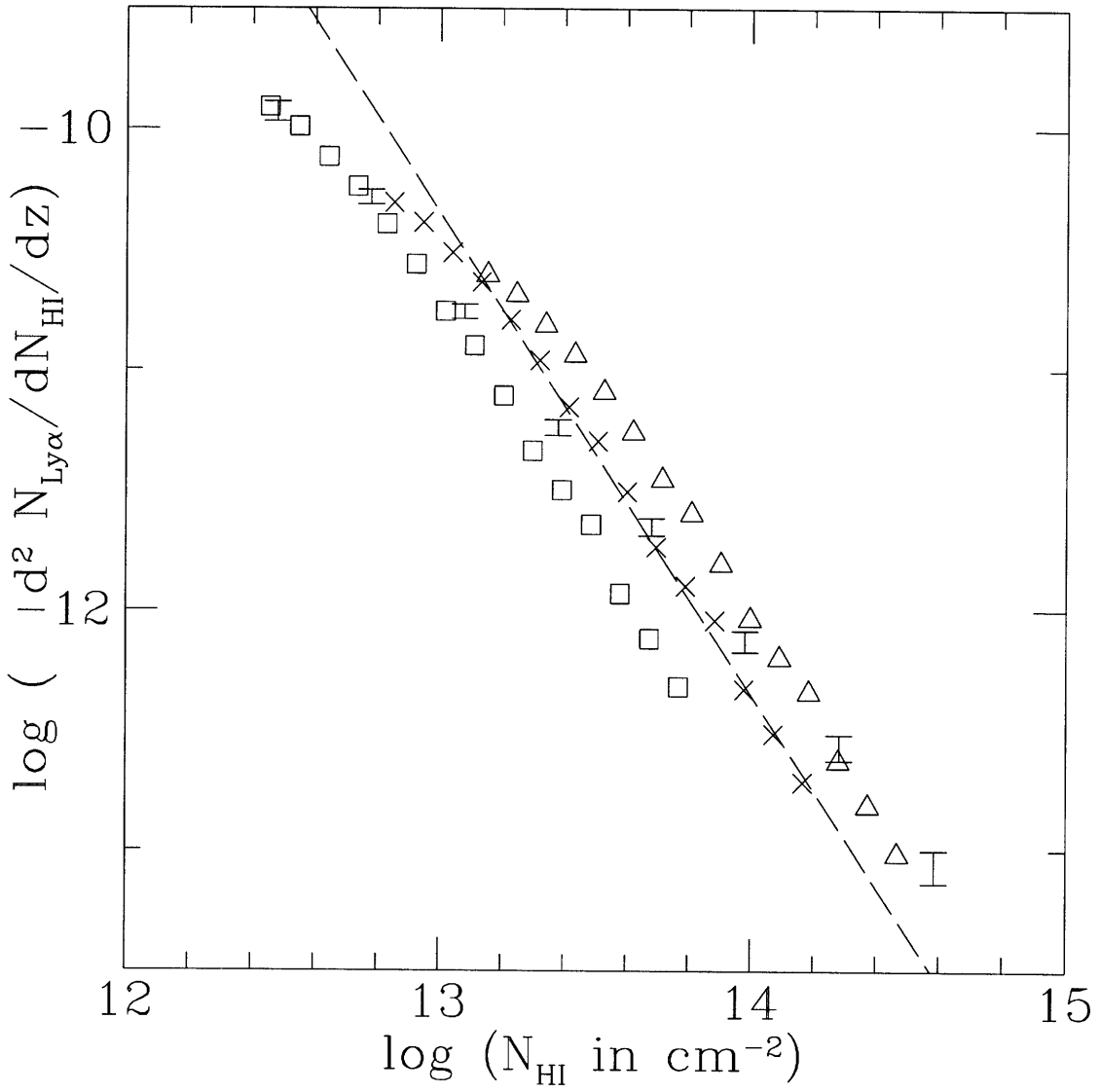

Fig. 6-19.— The column density distribution for the $B2$ CHDM model (Table 6-2). Three values of F are shown (equation [6.41]): $F = 1$ (open squares), $F = 2.5$ (crosses) and $F = 5$ (open triangles). $\gamma = 1.5$ for all three (equation [6.28]). Points with error-bars are the observational data as in Fig. 6-9. The dashed line has a slope of -2.

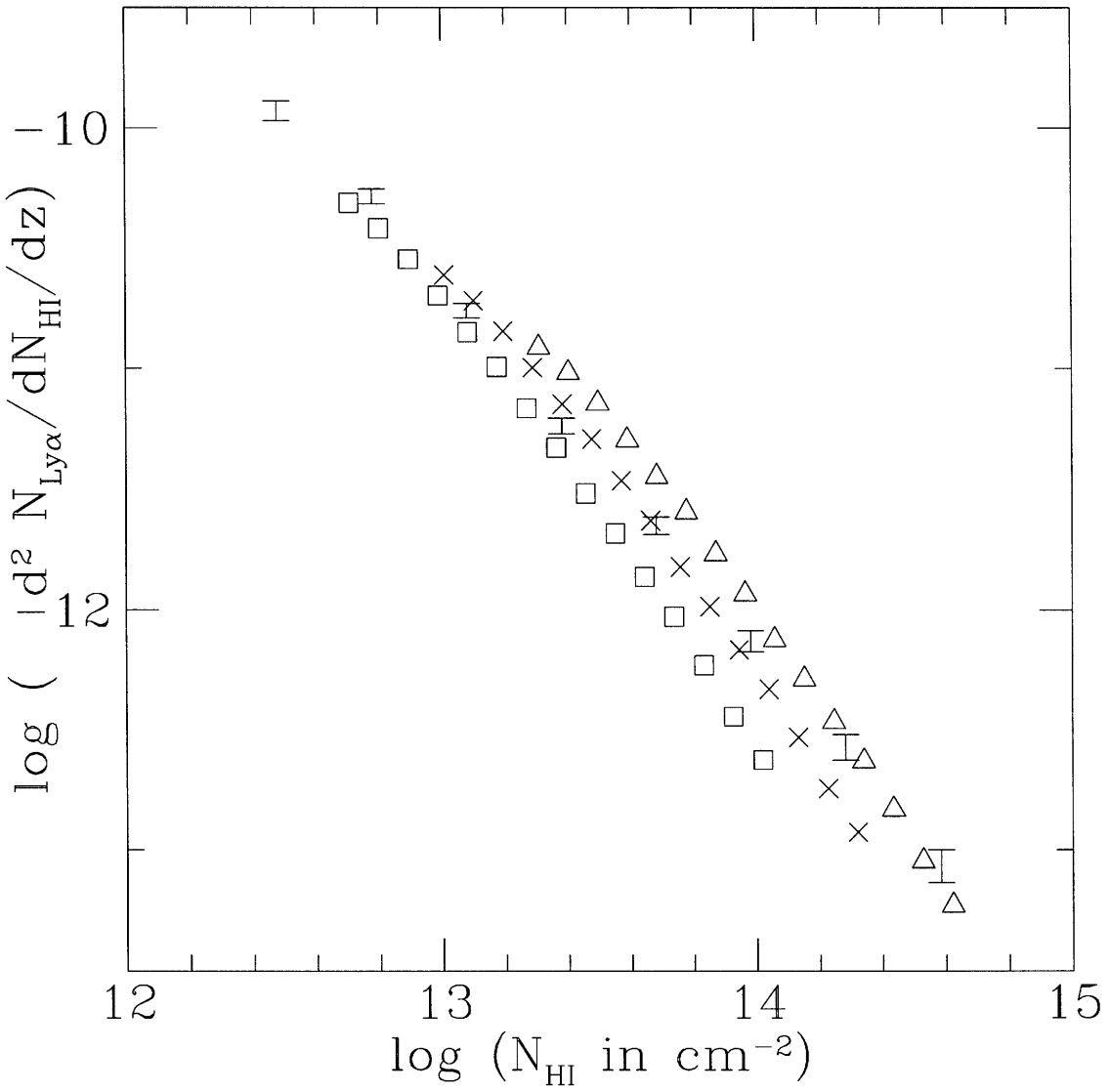

Fig. 6-20.— The column density distribution for the $D2$ (Table 6-2) CHDM model. Three values of F are shown (equation [6.41]): $F = 7.14$ (open triangles), $F = 3.57$ (crosses) and $F = 1.79$ (open squares). $\gamma = 1.5$ for all three (equation [6.28]). Points with error-bars are the observational data as in Fig. 6-9.

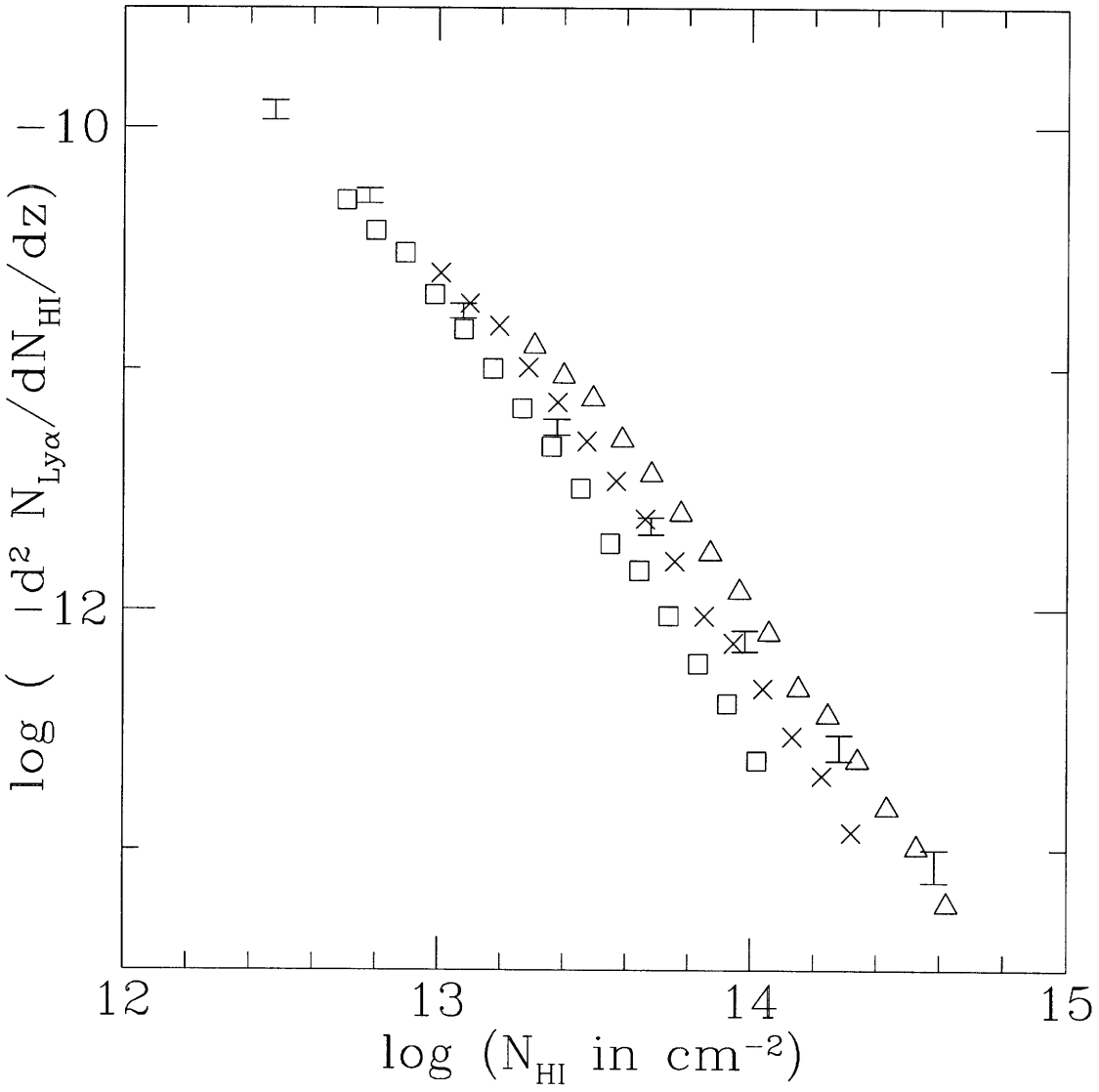

Fig. 6-21.— The column density distribution for the $C2$ CHDM model (Table 6-2). Three values of F (equation [6.41]) are shown: $F = 7.14$ (open squares), $F = 3.57$ (crosses) and $F = 1.79$ (open triangles). $\gamma = 1.5$ for all three (equation [6.28]). Points with error-bars are the same observational data as in Fig. 6-9.

than those with $\Omega_\nu = 0.1$ on scales $1\mathrm{Mpc}^{-1} < k_S < 10\mathrm{Mpc}^{-1}$, which are relevant for the range of column densities we are interested in. As we have explained before, the column density distributions are therefore steeper for the $\Omega_\nu = 0.2$ models at this range of column densities because they have relatively higher number of low density peaks compared to high density peaks. Among the $\Omega_\nu = 0.2$ models, those with a lower Hubble constant produce comparatively steeper column density distributions because they have even less small scale power than the ones with a higher Hubble constant.

In Sec. 6.6, we have discussed how the equation of state can change the slope of the column density distribution although the effect is very small. To demonstrate the robustness of our conclusion, we show in Fig. 6-22 the effects of altering the equation of state on the column density distribution for one particular CHDM model ($A2$). F is fixed at 2.5, the value that seems to give a column density distribution closest to the observational data. Smaller γ, as we have noted before, helps flatten the column density distribution but the flattening seems to be not quite enough even for $\gamma = 1.2$. We show in the same figure a dashed line with a slope of -1.77 (which follows from equation (6.45) if one fixes m using the case of $\gamma = 1.5$ and $\beta = 2$ in Fig. 6-18). It seems $\gamma < 1.2$ is needed for this model to give the right slope of the distribution, at least the right slope to within the 95% confidence limits of the observed value (-1.37,-1.51). The same conclusion holds for the other low Hubble constant $\Omega_\nu = 0.2$ model ($B2$). We should emphasize, however, that a more detailed comparison between the predictions of the models and observations, taking into account noise and biases of the line identification techniques, is necessary before any model can be considered ruled out.

The high Hubble constant ($h_{100} = 0.65$) $\Omega_\nu = 0.2$ models $C2$ and $D2$, on the other hand, have intrinsically flatter distributions and a reasonable match between theory and observations can be made by choosing γ in the range $1.2 - 1.7$. It remains true that this class of models have steeper column density distributions compared to the CDM models previously considered for the same value of γ.

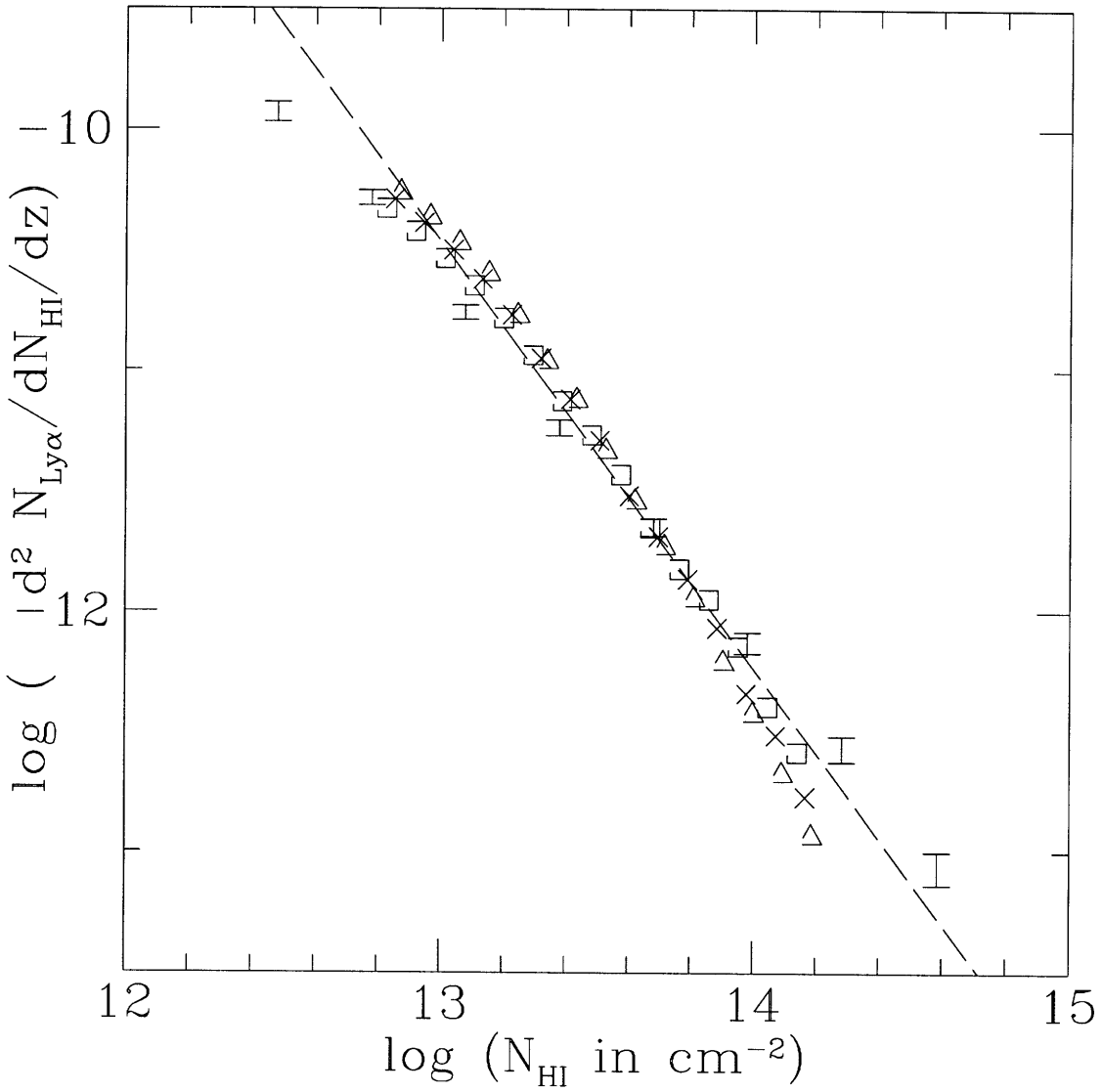

Fig. 6-22.— The column density distribution of the $A2$ CHDM model (Table 6-2) for three different values of γ (equation [6.28]). $F = 2.5$ (equation [6.41]) is adopted. Three values of γ are shown: $\gamma = 1.2$ (open squares), $\gamma = 1.5$ (crosses) and $\gamma = 1.7$ (open triangles). Points with error-bars are the observational data as in Fig. 6-9. The dashed line has a slope of -1.77 .

6.9 Conclusion

We have systematically developed a set of tools to compute in an efficient manner the column density distribution given a cosmological model. One fundamental assumption of the approximations involved is that most of the Lyman-α forest with column densities in the range $10^{12.5} - 10^{14.5}\,\mathrm{cm}^{-2}$ originate from regions of low overdensities or even underdensities which have not undergone orbit-crossing. The result of a comparison with a hydrodynamic simulation lends support to it.

One major conclusion we reach, in the process of developing the tools, is that the peculiar velocities play almost no role in determining the column density distribution at our column densities of interest, even though they are very important in determining the shapes of individual absorption line profiles. We take advantage of this fact and develop a method we call the Density-Peak-Ansätz in which each density peak is assigned a column density. The column density distribution then becomes a statistic of density peaks.

The main insight we gain in the two sections on the CDM and CHDM models is that the factors controlling the column density distribution can be divided into two categories. One mostly affects its normalization while the other mostly influences its slope. Those that fall into the former category include the ionizing radiation intensity, the mean temperature of the intergalactic medium and the mean baryon density. Uncertainties in their values are such that almost any viable cosmological model which have the correct slope of the column density distribution can be made to match observations by a judicious choice of parameters.

The factors that mostly affect the slope of the distribution include the equation of state and more strongly so, the amount of (linear) power on scales $1\,\mathrm{Mpc}^{-1} \lesssim k \lesssim 10\,\mathrm{Mpc}^{-1}$. Models that have less power on these scales tend to have comparatively more low density peaks to high density ones and hence have relatively steeper column density distributions. Equations of state which are closer to isothermal (smaller γ where γ satisfies $T \propto (1+\delta_\mathrm{b})^{\gamma-1}$) tend to produce flatter column density distributions. However, within the reasonable range of γ (see Hui and Gnedin 1996), its precise value

depending upon the reionization history, the effect of changing the equation of state is small. We put forward approximation expressions relating the slope of the column density distribution to γ and the power on small scales (equation [6.45] is the more general expression and equation [6.47] holds for a class of CDM models.)

Therefore, the slope of the column density distribution provides a rather robust measure of the amount of power on small scales for a given cosmological model. We apply our techniques to study a class of CHDM models which are known to have less power on small scales compared to other popular CDM models. We conclude that the CHDM models indeed produce steeper column density distributions compared to the CDM models. In particular, the low Hubble constant ($h_{100} = 0.5$) $\Omega_\nu = 0.2$ CHDM models, which have the least amount of power on small scales among the models we study, have column density distributions which can be made consistent with observations only for γ less than the values we consider reasonable. We emphasize however that only after a more detailed comparison between theories and observations, including all the effects of noise and biases of the line-identification methods, can any model be considered ruled out, based on the observed column density distribution.

We therefore conclude that a lot of work still needs to be done both on the observational and theoretical fronts. The biases of the line-identification techniques used for data reduction deserve close study so that the error bars in the observed column density distributions can be better understood and perhaps reduced. Numerical simulations on the CHDM models should be carried out to test the accuracy of the approximations made in the present work. The effect of a fluctuating radiation field, instead of a uniform one as is assumed here, has to be investigated. Moreover, in terms of constraining models, it is also important to examine other possible statistics. We have shown, for instance, that the column density distribution is relatively independent of peculiar velocities. Are there other statistics that can take advantage of the different peculiar velocity structures predicted by different cosmological models? One question to consider is how the peculiar velocity affects the observed $N_{\rm HI} - b$ relation.

In short, the study of the Lyman-α forest has entered an exciting stage. There

is a gold mine of information contained in the quasar absorption spectra waiting to be discovered.

Chapter 7

Final Conclusion

A brief summary of all the main results are presented here. For more detailed discussions, the reader is referred to the separate conclusion to each of the three parts of the thesis in Chapters 3, 5 and 6.

The evolution of density fluctuations in the inflationary universe is re-examined in Chapters 2 and 3. The transition between the inflaton-dominated and the radiation-dominated stages is approximated as instantaneous. If the transition occurs during the slow-roll phase of inflation, the widely used formula for the amplitude of density fluctuations at second Hubble-radius-crossing, $\delta\rho(k)/\rho_0 \propto H\delta\varphi(k)/\dot{\varphi}_0$, is recovered. However, if the transition takes place very close to the instant of turn-around, when the inflaton has rolled past the minimum of the potential, decelerated and comes to a halt momentarily, the above formula breaks down. The amplitude of density fluctuations is greatly enhanced. This is true even for fluctuations with wavelengths much longer than the Hubble radius at the time of the transition. This demonstrates an exception to the conventional wisdom that the details of the transition or reheating process cannot affect the amplitude of large scale density fluctuations.

Conventional wisdom has it that one can choose appropriate gauge invariant quantities, such as ζ defined in equations (2.14b) and (2.15), which remain constant as long as the wavelength of the fluctuation exceeds the Hubble radius. This leads to the commonly accepted picture is that once the wavelength of a given fluctuation is stretched beyond the Hubble radius by inflation, it (or the quantity describing it) is

frozen until at some point during the radiation-dominated or matter-dominated era, it enters the Hubble radius again.

The standard lore breaks down because ζ, or other analogous quantities, does change close to the point of turn-around. In fact, it diverges at exactly that point. However, this does not imply any physical divergence unless inflation ends right at the moment of turn-around. We have shown in Section 3.3 that Φ, which is proportional to a gauge invariant density fluctuation, remains finite throughout the turn-around. It is the divergence of ζ at that particular moment, together with the role it plays in the junction conditions, that causes the physical divergence of density fluctuations at second Hubble-radius crossing, if one forces inflation to end precisely at the moment of turn-around.

A highly idealized model of instantaneous transition is employed in the calculation presented in this thesis. In reality, the transition probably takes at least a few oscillations of the inflaton around the true minimum of the potential to complete. Since the departure from standard results is found to occur only if inflation is abruptly ended extremely close to the point of turn-around, it is quite possible that the effect we discover is not important in a more realistic calculation. Moreover, since divergence of ζ at precisely the turn-around occurs because the sum of energy density and pressure is strictly zero at that point, any small amount of matter or radiation that is present during the point of turn-around can kill the divergence. Nevertheless, it remains to be shown under what circumstances reheating does/does not affect the large scale density fluctuations. The calculation presented provides an intriguing example of what can happen.

In Chapters 4 and 5, the concept of local approximations is introduced and three examples are discussed. It is shown how the celebrated Zel'dovich approximation can be viewed as a member of a larger class of approximations which involve the truncation of a hierarchy of Lagrangian fluid equations. The purpose of the truncation is to produce a closed set of equations describing the evolution of the density, velocity gradients and tidal field at a mass element without coupling to other mass elements.

Two approximations based on ignoring certain terms in the tidal evolution equation are discussed. It is shown that one of them, the Local Tidal approximation (LTA), is exact for cases of planar, cylindrical and spherical symmetries. This is a significant improvement over the Zel'dovich approximation which is exact only in the case of planar symmetry. The draw-back is that we have a more complicated set of equations to solve which have to be handled numerically except under special circumstances. The local approximations are compared in the case of ellipsoidal collapse. It is shown that the Local Tidal approximation gives a reasonably accurate estimate of the time of pancake collapse while the Zel'dovich approximation overestimates it. The third local approximation, known as the Non-magnetic approximation (NMA), gives rise to collapse into the spindle configuration, contrary to what actually happens. It shows that local approximations in the language of truncated Lagrangian fluid equations have to be formulated with care. We show which terms in the tidal evolution equation are responsible for encouraging pancake versus spindle collapse.

The successful local approximations can be applied in the mildly nonlinear regime (overdensity less than about 5) before trajectory crossing takes place. An application of the Zel'dovich approximation to the study of the Lyman-α forest is presented in Chapter 6.

The focus of the last chapter is on understanding the factors that influence the column density distribution of the Lyman-α forest. To that end, we first show how the Zel'dovich approximation, with appropriate smoothing and coupled with the thermal and ionization evolution equations, can be employed to compute the neutral hydrogen density distribution as well as the peculiar velocity field, for any cosmological model. This is proposed as an efficient and accurate alternative to full hydrodynamic simulations for the part of the Lyman-α forest that has column density in the range $10^{12.5} - 10^{14.5}\mathrm{cm}^{-2}$. The point of view is that this part of the forest originates from fluctuations of mild overdensities or even underdensities which arise naturally in hierarchical clustering models.

Then we develop a method we call the Density-Peak-Ansätz to evaluate the column density distribution. It consists of two parts: first, each density peak in real

space is assumed to give rise to an absorption line; second, each peak is then assigned a column density based its height and its second derivative along the line of sight. The first step is justified by the fact the peculiar velocities play a relatively minor role in determining the column density distribution even though it is important in shaping absorption profiles. The Density-Peak-Ansätz is tested against the result of a full hydrodynamic simulation where the column density distribution is obtained by analyzing the simulated spectra in a way similar to what observers do. The level of agreement is excellent. Using the ansätz, the column density distribution becomes a statistic of density peaks in real space that involves its height and second derivative.

The method is systematically applied to study how the column density distribution varies with the power spectrum and various properties of the intergalactic medium. It is found that the column density distribution can indeed be approximated as a power law in our limited range of investigation, for the cosmological models we study. It is shown that the baryon density, ionization flux and temperature of the medium largely affects the normalization of the column density distribution. The equation of state, on the other hand, mostly affects its slope although the effect is found to be small for a reasonable range of equations of state. Lastly, it is found that the column density distribution is sensitive to the amount of power on scales $1\,\mathrm{Mpc}^{-1} \lesssim k \lesssim 10\,\mathrm{Mpc}^{-1}$. Models that have less power on those scale tend to have more density peaks of intermediate densities compared to high densities. The implication is that the column density distributions of the models with less power are steeper. At sufficiently low column densities however, the reverse should be true because the very low column density objects are dominated by very low density peaks which are rare in models with less power. For the range of column densities that we can reliably compute the distribution, $10^{12.5} - 10^{14.5}\mathrm{cm}^{-2}$, we can see the former effect but not the later. Presumably, the latter holds at column densities below our resolution limit. The steepening effect is demonstrated by varying the normalization and the spectral index of the Cold Dark Matter (CDM) model. An analytical expression is developed to describe the dependence of the slope of the column density distribution on the equation of state as well as the amount of power on small scales.

We study a class of Cold plus Hot Dark Matter (CHDM) models with varying neutrino content and Hubble constant and confirm the above trend. Our comparison with the observational data show that the low Hubble constant ($H_0 = 50\,\mathrm{kms}^{-1}\,\mathrm{Mpc}^{-1}$) $\Omega_\nu = 0.2$ CHDM models produce column density distributions that are too steep unless the equation of state is chosen to be close to isothermal ($\gamma < 2$ where γ satisfies $T \propto (1+\delta_b)^{\gamma-1}$; T is the temperature and δ_b is the overdensity). It is emphasized that no cosmological model should be considered ruled out until a more detailed comparison with the observational data is carried out, which should take into account the effects of noise and biases of line-identification methods in analyzing both the observed and simulated data. Preferably the two should be analyzed in exactly the same way but the approximate methods we introduce to compute the theoretical column density distribution are useful for gaining insights into the factors that determine it. Its quantitative success when compared to one hydrodynamic simulation encourages us to apply it to study the CHDM models. The general trend of steeper a column density distribution in the range $10^{12.5} - 10^{14.5}\mathrm{cm}^{-2}$ with higher neutrino content and smaller Hubble constant has important implications for the CHDM models.

Appendix A

Relativistic Perturbation Theory

Our treatment of gauge invariant metric variables and their evolution follows closely that of Bardeen (1980). The symbols for metric and energy-momentum-tensor perturbations are defined in equations (2.11) and (2.12). Assume the following coordinate transformation:

$$\tilde{\eta} = \eta + N(\eta)Q(\mathbf{x}) \tag{A.1a}$$

$$\tilde{x}^i = x^i + L(\eta)\frac{Q^{,i}(\mathbf{x})}{\mathbf{k}} \tag{A.1b}$$

where Q is defined in equation (2.10).

Then the metric variables transform as follows:

$$\tilde{A} = A - N' - \frac{a'}{a}N \tag{A.2a}$$

$$\tilde{B} = B + L' + kN \tag{A.2b}$$

$$\tilde{H}_L = H_L - \frac{k}{3}L - \frac{a'}{a}N \tag{A.2c}$$

$$\tilde{H}_T = H_T + kL \tag{A.2d}$$

The components of the energy-momentum tensor transform as:

$$\tilde{v} = v + L' \tag{A.3a}$$

$$\tilde{\delta\rho} = \delta\rho - \rho_0' N \tag{A.3b}$$

$$\tilde{\delta p} = \delta p - p_0' N \tag{A.3c}$$

The anisotropic stress, which is the traceless part of T^i_j, is gauge invariant to first order. Hence if it vanishes in one coordinate system, it vanishes in all the others.

It is simple to show from the above transformations that the quantities $\delta\rho_{GI}$ and δp_{GI} (equation [2.13]) are gauge invariant. The gauge invariance of Φ, Φ_p and ζ then follows straight from their definitions. Since $\delta\varphi$ transforms like

$$\tilde{\delta\varphi} = \delta\varphi - \varphi_0' N\,, \tag{A.4}$$

$\delta\varphi_{GI}$ (equation [2.26]) is also gauge invariant.

Next we derive the key equation used in Chapters 2 and 3, (2.17). The first order Einstein Equations are:

$$\delta\rho = \frac{3}{a^2}\left[2\frac{a'}{a}H_L' + \frac{2k^2}{3}(H_L + \frac{1}{3}H_T) + \frac{2k}{3}\frac{a'}{a}B - 2\left(\frac{a'}{a}\right)^2 A\right] \tag{A.5a}$$

$$(\rho_0 + p_0)(v - B) = \frac{2k}{a^2}\left(-H_L' - \frac{1}{3}H_T' + \frac{a'}{a}A\right) \tag{A.5b}$$

$$H_T'' + 2\frac{a'}{a}H_T' - k\left(B' + 2\frac{a'}{a}B\right) - k^2\left(H_L + \frac{1}{3}H_T + A\right) = 0 \tag{A.5c}$$

The first two equations follow from the time-time and time-space components of the Einstein Equations. The last equation is an expression of the no-anisotropic-stress assumption. We do not list the equation associated with the trace of T^i_j or the pressure. It is not needed in the following derivation. However, we need the energy and momentum conservation equations, which can be derived either from the above equations together with the pressure equation, or more simply, from $T^{\mu}_{\nu\ ;\mu} = 0$.

$$(a^2\delta\rho)' = -\frac{a'}{a}(a^2\delta\rho) - a^2(\rho_0 + p_0)(kv + 3H_L') - 3\frac{a'}{a}(a^2\delta p) \tag{A.6a}$$

$$(v - B)' = -\frac{a'}{a}(v - B) + kA + \frac{k}{\rho_0 + p_0}\delta p_{GI}\,. \tag{A.6b}$$

δp_{GI} is defined in equation (2.13).

Putting equations (A.5a) and (A.5b) into equation (2.13a) and using the definition of Φ in equation (2.14a), we obtain:

$$k^2\Phi = -\frac{a^2}{2}\delta\rho_{GI} \tag{A.7}$$

This resembles the Poisson equation.

Combining the two conservation equations and using the above expression, it can be shown that:

$$\Phi' = -\frac{a'}{a}\Phi + \frac{a^2}{2k}(\rho_0 + p_0)\left(v - \frac{H_T'}{k}\right) . \tag{A.8}$$

Equation (A.5b) is also used in the derivation.

The above expression is differentiated once. It can be shown that, using equations (A.6b) and (2.6) to simplify,

$$\Phi'' - \frac{(\rho_0 + p_0)'}{\rho_0 + p_0}\Phi' - \frac{a'}{a}\frac{(\rho_0 + p_0)'}{\rho_0 + p_0}\Phi = a^2(\rho_0 + p_0)\Phi + \frac{a^2}{2}\delta p_{GI} \tag{A.9}$$

Equation (2.17) follows if we replace δp_{GI} using equation (2.18) and change the independent variable from conformal to proper time. The key equation of Chapters 2 and 3 is verified.

Let us turn to the two definitions of ζ in equations (2.14b) and (2.15). The quantity $H_{\rm local}$ is defined as $\nabla_\alpha U^\alpha$, where U^α is the fluid velocity. The components are $U^0 = (1 - AQ)/a$ and $U^i = -vQ^{,i}/(ak)$. The covariant divergence of U^α divided by 3 is:

$$H_{\rm local} = \frac{a'}{a^2} - \frac{a'}{a^2}AQ + \frac{k}{3a}vQ + \frac{H_L'}{a}Q \tag{A.10}$$

The term $\rho/3$ in equation (2.15) can be broken into two parts $\rho_0/3 + \delta\rho Q/3$. The first part cancels the first term on the right hand side of equation (A.10). The second part (rewritten using equation [A.5a]), together with the expression for $H_{\rm local}$

above, can be substituted into the definition for ζ in equation (2.15) to give:

$$\zeta = -(H_L + \frac{1}{3}H_T) + \frac{1}{\gamma a H}(-H_L' - \frac{1}{3}H_T' + \frac{a'}{a}A) \tag{A.11}$$

where we have dropped the spatial harmonics Q. This expression can be shown to be consistent with the other definition for ζ (equation [2.14b]) by making use of equation (2.14a).

Finally we derive equation (2.27), which relates $\delta\varphi_{GI}$ to the variable ζ. Combining equations (2.25) and (A.5b), we obtain

$$\delta\varphi = \frac{2}{\varphi_0'}\left(-H_L' - \frac{1}{3}H_T' + \frac{a'}{a}A\right) \tag{A.12}$$

Substituting this into equation (2.26) and putting (2.14a) and (A.8) into equation (2.14b), $\zeta = aH\delta\varphi_{GI}/\varphi_0'$ can be verified.

Appendix B

The Peculiar Behavior of ζ

We have argued in the Chapter 2 that ζ is like a curvature fluctuation. We can make this precise by taking the small k limit of equation (2.17). The result should be that ζ is simply a constant because it is proportional to the curvature. It describes a homogeneous perturbation of a flat background. Putting $k = 0$ in equation (2.17) and using the definition of ζ in equation (2.14a), it can be shown that

$$\frac{d\zeta}{dt} = 0 \tag{B.1}$$

Note that in the above derivation, no assumption about the equation of state has been made.

Here is the puzzle: if the $k = 0$ limit of ζ does correspond to a homogeneous curvature fluctuation and hence is a constant, how can it possibly vary close to the point of turn-around, not to mention changing sign in the process (equation [3.9a])?

The answer is that by taking the $k = 0$ limit in the manner described above, one inevitably misses the second mode of ζ. For concreteness, let us think about the behavior of Φ and ζ in the radiation-dominated stage (because everything can be solved analytically).

Equation (2.33) is the counter-part of equation (2.17) in the case of the radiation-dominated universe. It can be seen that indeed, if one puts $k = 0$ in the equation, equation (B.1) can be obtained (see also equation [2.34]).

Let us follow a different strategy. Suppose k is non-zero, we can solve equation (2.33) and obtain the solution in equation (2.37). Then, take the small k limit of the solution. To be precise, we take the limit of $k\tau$ approaching zero. $k\tau$ is equal to k/aH which is the ratio of the proper wavelength to the Hubble radius. The lowest order term involving K_{r1} (equation [2.37]) is $2kK_{r1}/(3\sqrt{3})$. The lowest order term involving K_{r2} is $6kK_{r2}/(k\tau)^3$. Putting these into the definition of ζ in terms of Φ (equation [2.14b]), one obtains $kK_{r1}/\sqrt{3}$. This is the curvature mode we have been talking about. It is simply a constant. The mode involving the constant K_{r2} somehow cancels itself out of existence.

However, there is a higher order term involving K_{r2} in equation (2.37), which cannot be ignored if one takes the vanishing $k\tau$ limit. It is $kK_{r2}/(k\tau)$. Putting this into the definition of ζ in equation (2.14b), one obtains the other mode of ζ in the small $k\tau$ limit: $kK_{r2}/(k\tau)$. Putting the two modes of ζ together, we have

$$\zeta \sim \frac{kK_{r1}}{\sqrt{3}} + \frac{kK_{r2}}{k\tau} \tag{B.2}$$

The first mode on the right is the curvature mode whereas the second mode is the one we missed by simply putting $k = 0$ in the differential equation obeyed by Φ. Notice that it is not a constant. Alternatively, the same two modes can be obtained by putting $k = 0$ in the second order differential equation obeyed by ζ (equation [2.35], not the first order differential equation [2.34]).

The same reasoning can be carried over to the inflaton-dominated stage. The relevant equations in this case (equations [2.28] and [2.32]) cannot be solved explicitly except for special cases (for instance, if γ = constant which holds for power-law inflation). However, as in the case of the radiation-dominated stage, if one puts $k = 0$ in the second order differential equation [2.32]) (not the first order equation [2.29]), both modes of ζ are recovered, as in equation (3.2). One of them is the constant curvature mode, which is often used to argue for the freezing-out of the amplitude of density fluctuations at super-Hubble-radius scales. The other is responsible for the divergence at turn-around.

It is perhaps still a little puzzling why taking the small wavelength limit in the two different ways mentioned above should give us different results. Here is an alternative way to put it. Consider two different limits in the radiation dominated regime: one where k approaches zero at a fixed time τ while keeping $\Phi(k)$ fixed; the other where τ approaches zero at a fixed k (without keeping $\Phi(k)$ fixed by hand, but allowing it to evolve according to the evolution equation it obeys). The former corresponds to taking a fluctuation of a fixed amplitude and mentally placing it at bigger and bigger scale until the universe locally looks just like a homogeneous universe. The latter corresponds to focusing one's attention on a fluctuation of a fixed comoving wave number k and following its physical behavior backward in time.

The relevant equation for our discussion is equation (2.34), which can be rewritten as:

$$\frac{d\zeta}{d\tau} = -\frac{1}{6}k^2\tau\Phi \tag{B.3}$$

where we have used $aH = 1/\tau$ and $p_0 = \rho_0/3$, appropriate for the radiation dominated universe.

If one keeps Φ and τ fixed and take the vanishing k limit, it is obvious from the above equation that ζ approaches a constant i.e. the curvature mode.

However, let us examine again the following expansion of the solution for Φ from equation (2.37):

$$\Phi = \frac{2}{3\sqrt{3}}K_{r1}k + 6\frac{K_{r2}}{k^2\tau^3} + \frac{K_{r2}}{\tau} \tag{B.4}$$

where I have set $a = \tau$. All the terms that are ignored vanish in both of the limits under consideration.

For the first limit, keeping Φ and τ fixed and letting k go to zero means keeping $K_{r1}k$ and K_{r2}/k^2 constant (note that K_{r1} and K_{r2} are in general functions of k). Notice how in this limit the last term on the right hand side does not contribute to Φ at all. In any case, none of these terms survive on the right hand side of equation (B.3) once the factor of k^2 is taken to zero.

All these agree very well with what one would expect for a homogeneous perturbation. ζ has only one mode: a constant.

Consider the second limit. Putting equation (B.4) into equation (B.3), one obtains

$$\frac{d\zeta}{d\tau} = -\frac{1}{9\sqrt{3}} K_{r1} k^3 \tau - \frac{K_{r2}}{\tau^2} - \frac{1}{6} K_{r2} k^2 \tag{B.5}$$

Fix k and let τ approach zero. One can see that the derivative of ζ does not vanish anymore unless K_{r2} is chosen to vanish. In fact, ζ' diverges in the limit! This should really come as no surprise, it is simply the decaying mode in the radiation dominated universe, which blows up when it is extrapolated back in time.

The second limit is the appropriate one to consider for our application. After all, we are interested in the physical evolution of each wave mode in our calculation. The analog of the second limit for the inflationary era is taken when we derive the approximate solution for ζ in equation (3.2). It is obtained by ignoring k^2 in equation (2.32). This might sound dangerously close to the first limit discussed above. However, what we are ignoring is really k^2 compared to $(a\sqrt{\gamma})''/a\sqrt{\gamma}$ in equation (2.32). We have shown in Section 3.2 that this corresponds to taking the limit where the wavelength of interest is long compared to the Hubble radius in the slow-roll phase. Hence, the limit used in obtaining equation (3.2) does correspond to focusing on a wave mode of a particular k and watching how the fluctuation evolves as it exits the Hubble radius, only now as time going forward instead of going backward like in the radiation dominated case.

Appendix C

Second Order Calculation of M_{ij}

We write the Eulerian density fluctuation field $\delta = \delta^{(1)} + \delta^{(2)} + \ldots$ where $\delta^{(n)}$ is treated as being of nth order in perturbation theory. Similar expansions are used for the velocity field and the scaled gravitational potential

$$\hat{\phi} \equiv \frac{\phi}{4\pi G a^2 \bar{\rho}} = -\frac{1}{4\pi} \int d^3x' \, \frac{\delta(\vec{x}')}{|\vec{x} - \vec{x}'|} \, . \tag{C.1}$$

For simplicity we shall assume an Einstein-de Sitter universe as in the numerical examples presented. In this case the perturbation series is a series in $a(\tau)$.

Peebles (1980) presents the result for $\delta^{(2)}$, which we rewrite using our variables as

$$\delta^{(2)} = \frac{5}{7} \left[\delta^{(1)}\right]^2 + \left[\vec{\nabla}\delta^{(1)}\right] \cdot \left[\vec{\nabla}\hat{\phi}^{(1)}\right] + \frac{2}{7} F^2 \, , \tag{C.2}$$

where $F^2 \equiv F^{ij} F_{ij}$ and

$$F_{ij} \equiv \nabla_i \nabla_j \hat{\phi}^{(1)} \, ; \tag{C.3}$$

note that $F^i{}_i = \delta^{(1)}$. From the Euler equation (4.2) we get the first and second order terms of the peculiar velocity,

$$\vec{v}^{(1)} = -\frac{a'}{a} \vec{\nabla}\hat{\phi}^{(1)} \, , \quad \vec{v}^{(2)} = -\frac{2}{5}\frac{a'}{a} \left[\vec{\nabla}\hat{\phi}^{(1)} \cdot \vec{\nabla}\right] \vec{\nabla}\hat{\phi}^{(1)} - \frac{3}{5}\frac{a'}{a} \vec{\nabla}\hat{\phi}^{(2)} \, , \tag{C.4}$$

where $\hat{\phi}^{(2)}$ is obtained using equation (C.1) with $\delta^{(2)}$.

We get M_{ij} and $-\nabla_k \epsilon^{kl}{}_{(i}H_{j)l}$ from equation (4.23). They vanish in first order; the second order results are

$$M_{ij} = 4\pi G\bar{\rho}aa' \left\{ \delta^{(1)}F_{ij} - \frac{7}{5}\nabla_i\nabla_j\hat{\phi}^{(2)} + \frac{7}{5}\left[\vec{\nabla}\hat{\phi}^{(1)}\cdot\vec{\nabla}\right]F_{ij} + \frac{2}{5}F^k{}_{(i}F_{j)k} \right\} , \quad \text{(C.5)}$$

$$-\nabla_k \epsilon^{kl}{}_{(i}H_{j)l} = M_{ij} + 4\pi G\bar{\rho}aa' \left\{ 3\delta^{(1)}F_{ij} - 3F^k{}_{(i}F_{j)k} + \delta_{ij}\left(F^2 - \left[\delta^{(1)}\right]^2\right) \right\} . \quad \text{(C.6)}$$

It can be verified that these quantities are traceless as expected using equations (C.2) and (C.3). In general, neither vanishes in second order perturbation theory.

Appendix D

A List of Symbols

Tabulated below is a list of symbols used in the thesis. We try to identify equations in which or around which they are defined. In cases where it is impossible to do so, We list the equations where the meanings of the relevant symbols should be relatively clear. Their interpretation can also be found in the second column of the table below.

A word on the notation. Greek indices denote spacetime components and Latin indices are reserved for space components. c, the speed of light, is always set to one, unless stated otherwise (in Chapter 6). Factors of G are kept in Chapters 4 and 5 where the physics is Newtonian while $8\pi G$, as well as c, are set to one in Chapters 2 and 3.

There are a few cases where the same variables are used with different meanings in different places. For example, in Chapters 2 and 3 and Appendix A, the symbols α, β, σ and so on denote space-time indices while the same symbols represent quantities related to shear and tide in Chapter 5. We rely on the context to distinguish between the different meanings.

Table D-1: List of Symbols

Symbol	Meaning	Equation
a	Hubble scale factor	2.4
t	proper time	2.4
τ	conformal time/ optical depth	2.4 /6.1
ρ_0	average energy/mass density	2.5
p_0	average pressure	2.5
C	curvature	1.1
$\cdot$	derivative with respect to proper time	2.4
$'$	derivative with respect to conformal time	2.4
$H_{\rm infl}$	Hubble constant during inflation	1.6
t_i	starting proper time of inflation	1.6
φ_0	inflaton field	2.20
V	inflaton potential	2.21
ζ	gauge invariant curvature perturbation	2.14b and 2.15
Φ	gauge invariant potential perturbation	2.14a
Φ_p	gauge invariant pressure potential	2.18
$H_{\rm local}$	local Hubble parameter	2.15
H	Hubble parameter $\dot{a}/a$	
k	comoving wave-number	2.10
x	comoving spatial coordinate	
q	Lagrangian coordinate	4.4
v	peculiar velocity	1.11 & 2.12
δ	overdensity	1.10
$d/d\tau$	Lagrangian derivative	1.13
G	Newton's constant	1.1
$\delta\rho$	density fluctuation	2.12
δp	pressure fluctuation	2.12
$\delta\varphi_{GI}$	gauge invariant inflaton fluctuation	2.26
δp_{GI}	gauge invariant pressure fluctuation	2.13
$\delta\varphi$	inflaton fluctuation	2.20

Table D-1: List of Symbols (cont'd)

Symbol	Meaning	Equation
$T_{\mu\nu}$	energy-momentum tensor	2.12
$g_{\mu\nu}$	metric tensor	2.11
γ	$-\dot{H}/H^2$ or temperature-density index	2.8 / 6.28
c_s^2	p_0'/ρ_0'	2.9
Q	spatial harmonics	2.10
A	metric perturbation (time-time)	2.11
B	metric perturbation (time-sapce)	2.11
H_L	metric perturbation (space-sapce)	2.11
H_T	metric perturbation (space-sapce)	2.11
$h\,,h_\ell$	metric perturbations (synchr. gauge)	2.40
$K_{r1}\,,K_{r2}$	integration constants (rad. dom.)	2.36
$K_{q1}\,,K_{q2}$	integration constants (early inflation)	3.1
$K_1\,,K_2$	integration constants (late inflation)	3.2
$\text{subscript}_{\text{tr}}$	evaluation at transition: infl. to rad. dom.	2.43
$\text{subscript}_{\text{cr}}$	evaluation at 1st Hubble radius crossing	3.2
$\text{subscript}_{\text{v}}$	evaluation when valley phase starts	3.3
ζ_{sl}	asymptotic value of ζ during slow-roll	3.4
$\text{subscript}_{\text{infl}}$	evaluation on inflation side of transition	2.43
$\text{subscript}_{\text{rad}}$	evaluation on rad. dom. side of transition	2.43
x_{tr}	$k\tau_{\text{tr}}/\sqrt{3}$	2.43
$C(t)\,,D(t)$	coefficients of equation for ζ	3.5
$X(t)\,,Y(t)$	coefficients of equation for Φ	3.7
$F\,,G$	constants at the end of inflation	2.43
ζ_{infl}	ζ at the end of inflation	2.43
Φ_{infl}	Φ at the end of inflation	2.43
$\Delta(1/\gamma)$	Change in $1/\gamma$ across transition	2.41
$\delta\rho/\rho_0\vert_{\text{asymp}}$	amplitude of oscillation (rad. dom.)	2.45

Table D-1: List of Symbols (cont'd)

Symbol	Meaning	Equation
ψ^i	i-th component of displacement	4.4
ϕ	gravitational potential	4.3
ρ	full mass/energy density	4.6
$D_\pm$	linear growing and decaying modes	4.7
f	$d\ln D_+/d\ln a$	4.9
$\Omega 0$	ratio of density today to critical density	4.9
$H0$	Hubble parameter today	4.9
θ	expansion	4.10
σ_{ij}	shear tensor	4.10
ω_{ij}	vorticity tensor	4.10
δ_{ij}	Kronecker delta	4.10
ω_k	vorticity vector	4.10
ω^2	dot product of vorticity with itself	4.12
E_{ij}	gravitational tidal field	4.14
H_{ij}	magnetic part of the Weyl tensor (fluid threading)	4.19
$\vec{H}$	magnetic vector	4.20
$f_\perp$, $f_\parallel$	transverse and longitudinal parts of mass current	4.21
M_{ij}	combination of H_{ij} and E_{ij}	4.23
H_{ij}	magnetic part of the Weyl tensor (comoving threading)	4.27
R_i	length of the i-th axis of ellipsoid	5.1
ρ_b	average density of background universe	5.1
ρ_e	density of ellipsoid	5.1
α_i	i-thelliptic coefficient	5.2
X_i	initial axis ratio of ellipsoid	5.4
α_{i0}	intial i-th elliptic coefficient	5.4
δ_0	linear extrapolated overdensity at $a = 1$	5.4
ρ_c	density of cylinder	5.11
Q_{ij}	tidal angular matrix	5.13
α , β	tidal angular variables	5.12
σ , ϵ	tidal scalars	5.12
α_0 , ϵ_0	initial value of α and ϵ	5.16
a_c	Hubble scale factor at which collapse occurs	5.16

Table D-1: List of Symbols (cont'd)

Symbol	Meaning	Equation
σ_α	Lyman-α absorption cross-section	6.5
u	velocity coordinate	6.3
ν , ν_0 , ν_α	frequencies	6.3 and 6.4
b	b-value or thermal width	6.5
$n_{\rm H}$, $n_{\rm HI}$	total and neutral hydrogen density	6.8
$X_{\rm HI}$	neutral hydrogen fraction	6.9
$k_{\rm J}$	Jeans wavenumber	6.18
$k_{\rm NL}$	nonlinear scale	6.20
$k_{\rm S}$	truncation scale	6.21
$J_{\rm HI}$	radiation parameter	6.25
$\delta_{\rm b}$, $\delta_{\rm DM}$	overdensity in baryons and dark matter	6.8 and 6.10
ξ	logarithm of $1+\delta_b$	6.39
F	normalization factor of column density	6.41
σ_0	linear rms density fluctuation	6.43
β	- slope of column density distribution	6.45

Bibliography

Albrecht, A. & Steinhardt, P. J., 1982, *Phys. Rev. Lett.* **48**, 1220

Bagla, J. S. & Padmanabhan, T., 1994, *M.N.R.A.S.* **266**, 227

Bajtlik, S., 1993, in *The Environment and Evolution of the Galaxies*, ed. Shull, J. M., Thronson, H. A., Dordrecht: Kluwer

Bardeen, J. M., 1980, *Phys. Rev.* **D22**, 1882

Bardeen, J. M., Bond, J. R., Kaiser, N., & Szalay, A. S., 1986, *Ap.J.* **304**, 15

Bardeen, J. M., Steinhardt, P. J., & Turner, M. S., 1983, *Phys. Rev.* **D28**, 679

Barnes, A. & Rowlingson, R. R., 1994, *M.N.R.A.S.* **266**, 227

Barrow, J. D. & Saich, P., 1993, *Class. Quantum Gravity* **10**, 79

Batjlik, S., Duncan, R., & Ostriker, J. P., 1988, *Ap.J.* **327**, 570

Bechtold, J., 1994, *Ap.J. (Suppl)* **91**, 1

Bennett, C. L., Kogut, A., Hinshaw, G., Banday, A. J., Wright, E. L., Gorski, K. M., Wilkinson, D. T., Weiss, R., Smoot, G. F., Meyer, S. S., Mather, J. C., Lubin, P., Loewenstein, K., Lineweaver, C., Keegstra, P., Kaita, E., Jackson, P. D., & Cheng, E. S., 1994, *Ap.J.* **436**, 423

Bertschinger, E., 1995, *Cosmology and Large Scale Structure*, Proceedings of the Les Houches XV Summer School, in press

Bertschinger, E. & Hamilton, A. J. S., 1994, *Ap.J.* **435**, 1

Bertschinger, E. & Jain, B., 1994, *Ap.J.* **431**, 486

Bi, H., Ge, J., & Fang, L.-Z., 1995, *Ap.J.* **452**, 90

Bi, H. G., Borner, G., & Chu, Y., 1992, *Astronomy & Astrophysics* **266**, 1

Black, J. H., 1981, *M.N.R.A.S.* **197**, 553

Blau, S. K. & Guth, A. H., 1987, in *300 Years of Gravitation*, ed. Hawking, S. and

Israel, S., Cambridge University Press

Bond, J. R. & Myers, S. T., 1993, preprint

Bond, J. R., Szalay, A. S., & Silk, J., 1988, *Ap.J.* **324**, 627

Brainerd, T. G., Scherrer, R. J., & Villumsen, J. V., 1993, *Ap.J.* **418**, 570

Bruni, M., Matarrese, S., & Pantano, O., 1995, *Ap.J.* **445**, 958

Buchert, T., 1993, *Astronomy & Astrophysics* **267**, L51

Carswell, R. F., Lanzetta, K. M., Parnell, H. C., & Webb, J. K., 1991, *ApJ*, **371**, 36

Carswell, R. F., Webb, J. K., Baldwin, J. A., & Atwood, B., 1987, *Ap.J.* **319**, 709

Cen, R., Miralda-Escudé, J., Ostriker, J. P., & Rauch, M., 1994, *Ap.J. (Letters)* **437**, L9

Coles, P., Melott, A. L., & Shandarin, S. F., 1993, *M.N.R.A.S.* **260**, 765

Cristiani, S., D'Odorico, S., Fontana, A., Giallongo, E., & Savaglio, S., 1995, *MNRAS*, **273**, 1016

Deruelle, N. & Mukhanov, V. F., 1995, *Phys. Rev.* **D52**, 5549

Doroshkevich, A. G., 1970, *Astrofizika* **6**, 581

Duncan, R. C., Vishniac, E. T., & Ostriker, J. P., 1991, *Ap.J.* **368**, L1

Ehlers, J., 1961, *Akad. Wiss. Lit. Mainz Abh. Math.-Nat. Kl. 11*

Eisenstein, D. J. & Loeb, A., 1995, *Ap.J.* **439**, 520

Ellis, G. F. R., 1971, in *General Relativity and Cosmology*, ed. R. K. Sachs, New York: Academic

Fernandez-de Soto, A., Lanzetta, K. M., Barsons, X., Carswell, R. F., Webb, J. K., & Yahil, A., 1996, *Ap.J.* **460**, L61

Friedmann, A., 1922, *Z. Phys.* **10**, 377

Giallongo, E., Cristiani, S., D'Odorico, S., Fontana, A., & Savaglio, S., 1996, *ApJ*, **466**, 46+

Gnedin, N. Y. & Hui, L., 1996, preprint

Grishchuk, L. P., 1994, *Phys. Rev.* **D50**, 7154

Gunn, J. E. & Peterson, B. A., 1965, *Ap.J.* **142**, 1633

Guth, A. H., 1981, *Phys. Rev.* **D23**, 347

Guth, A. H. & Pi, S. Y., 1982, *Phys. Rev. Lett.* **49**, 1110

Guth, A. H. & Pi, S. Y., 1985, *Phys. Rev.* **32**, 1899

Hawking, S., 1966, *Ap.J.* **145**, 544

Hawking, S. W., 1982, *Phys. Lett.* **115B**, 295

Hernquist, L., Katz, N., Weinberg, D. H., & Miralda-Escudé, J., 1995, *Ap.J. (Letters)* **457**, L51

Hockney, R. W. & Eastwood, J. W., 1988, *Computer Simulation Using Particles*, Adam Hilger

Hogan, C. & Rugers, M., 1996, preprint, astro-ph 9603084

Hu, E., Kim, T., Cowie, L. L., Songaila, A., & Rauch, M., 1995, *A.J.* **110**, 1526

Hui, L. & Bertschinger, E., 1996, accepted for publication in *Ap.J.* November issue

Hui, L. & Gnedin, N. Y., 1996, in preparation

Hui, L., Gnedin, N. Y., & Zhang, Y., 1996, preprint

Hui, L. & Guth, A. H., 1996, preprint

Icke, V., 1973, *Astronomy & Astrophysics* **27**, 1

Jenkins, E. B. & Ostriker, J. P., 1991, *Ap.J.* **376**, 33

Kauffmann, G. & Charlot, S., 1994, *Ap.J. (Letters)* **430**, L97

Kofman, L., 1991, in *Primordial Nucleosynthesis and Evolution of Early Universe*, ed. K. Sato and J. Audouze, Dordrecht: Kluwer

Kofman, L., Bertschinger, E., Gelb, J. M., Nusser, A., & Dekel, A., 1994, *Ap.J.* **420**, 44

Kofman, L. & Pogosyan, D. Y., 1995, *Ap.J.* **442**, 30

Kofman, L., Pogosyan, D. Y., & Shandarin, S. F., 1990, *M.N.R.A.S.* **242**, 200

Kundt & Trumper, 1961, *Akad. Wiss. Lit. Mainz Abh. Math.-Nat. Kl.* p. 12

Lin, C. C., Mestel, L., & Shu, F. H., 1965, *Ap.J.* **142**, 1433

Linde, A. D., 1982, *Phys. Lett.* **108B**, 389

Linde, A. D., 1990, *Particle Physics and Inflationary Cosmology*, New York: Harwood

Lu, L., Sargent, W. L. W., Womble, D. S., & Takada-Hidai, M., 1991a, preprint, astro-ph 9605198

Lu, L., Wolfe, A. M., & Turnshek, D. A., 1991b, *Ap.J.* **367**, 19

Lynds, C. R., 1971, *Ap.J. (Letters)* **164**, L73

Ma, C.-P., 1996, preprint, astro-ph 9605198

Ma, C.-P. & Bertschinger, E., 1994, *Ap.J. (Letters)* **434**, L5

Matarrese, S., Lucchin, F., Moscardini, L., & Saez, D., 1992, *M.N.R.A.S.* **259**, 437

Matarrese, S., Pantano, O., & Saez, D., 1993, *Phys. Rev.* **D47**, 1311

Matarrese, S. & Terranova, D., 1995, preprint, astro-ph 9511093

McGill, C., 1990, *M.N.R.A.S.* **242**, 544

Meiksin, A. & Bouchet, F. R., 1995, *ApJLett*, **448**, L85

Melott, A. L., 1994, *Ap.J. (Letters)* **426**, L19

Melott, A. L., Buchert, T., & Weiss, A., 1995, *Astronomy & Astrophysics* **294**, 345

Melott, A. L., Sathyaprakash, B. S., & Sahni, V., 1996, *Ap.J.* **456**, 65

Miralda-Escudé, J., Cen, R., Ostriker, J. P., & Rauch, M., 1995, preprint, astro-ph 9511013

Mo, H. J. & Miralda-Escudé, J., 1994, *Ap.J. (Letters)* **430**, L25

Mukhanov, V. F., Feldman, H. A., & Brandenberger, R. H., 1992, *Phys. Report* **215**, 203

Murdoch, H. S., Hunstead, R. W., Petinni, M., & Blades, J. C., 1986, *Ap.J.* **309**, 19

Ostriker, J. P. & Ikeuchi, S., 1983, *Ap.J.* **268**, L63

Padmanabhan, T., 1993, *Structure Formation in the Universe*, Cambridge University Press

Pando, J. & Fang, L.-Z., 1996, *ApJ*, **459**, 1

Peebles, P. J. E., 1980, *The Large Scale Structure of the Universe*, Princeton University Press

Peebles, P. J. E., 1993, *Principles of Physical Cosmology*, Princeton University Press

Peterson, B. A., 1978, in *The Large Scale Structure of the Universe*, ed. Longair, M. S. and Einasto, J., Dordrecht: Reidel

Petitjean, P., Mücket, J. P., & Kates, R. E., 1995, *Astronomy & Astrophysics* **295**, L9

Petitjean, P., Webb, J. K., Rauch, M., Carswell, R. F., & Lanzetta, K., 1993, *M.N.R.A.S.* **262**, 499

Press, W. H. & Rybicki, G. B., 1993, *Ap.J.* **418**, 585

Press, W. H., Rybicki, G. B., & Schneider, D. P., 1993, *Ap.J.* **414**, 64

Press, W. H. & Schechter, P., 1974, **187**, 425

Ratra, B., 1991, *Phys. Rev.* **D43**, 3802

Rees, M., 1986, *M.N.R.A.S.* **218**, 25p

Reisenegger, A. & Miralda-Escudé, J., 1995, *Ap.J.* **449**, 476

Robertson, H. P., 1935, *Ap.J.* **83**, 257

Rybicki, G. B. & Lightman, A. P., 1979, *Radiative Processes in Astrophysics*, New York: Wiley

Sachs, R., 1961, *Proc. Roy. Soc. London* **264**, 309

Sargent, W. L. W., Young, P. J., Bokensberg, A., & Tytler, D., 1980, *Ap.J. (Suppl)* **42**, 41

Shandarin, S. F., Melott, A., McDavitt, K., Pauls, J. L., & Tinker, J., 1995, *Phys. Rev. Lett.* **75**, 7

Shandarin, S. F. & Zel'dovich, Y., 1989, *Rev. Mod. Phys.* **61**, 185

Smoot, G. F., Bennett, C. L., Kogut, A., Aymon, J., Backus, C., De Amici, G., Galuk, K., Jackson, P. D., Keegstra, P., Rokke, L., Tenorio, L., Torres, S., Gulkis, S., Hauser, M. G., Janssen, M. A., Mather, J. C., Weiss, R., Wilkinson, D. T., Wright, E. L., Boggess, N. W., Cheng, E. S., Kelsall, T., Lubin, P., Meyer, S., Moseley, S. H., Murdock, T. L., Shafer, R. A., & Silverberg, R. F., 1991, *Ap.J. (Letters)* **371**, L1

Songaila, A. & Cowie, L. L., 1996, preprint, astro-ph 9605102

Spitzer, L., 1978, *Physical Processes in the Interstellar Medium*, New York: Wiley

Starobinsky, A. A., 1982, *Phys. Lett.* **117B**, 175

Tytler, D., 1987, *Ap.J.* **321**, 49

Tytler, D., 1992, in *Statistical Challenges in Modern Astronomy*, ed. by Feigelson, E. D. and Babu, G. J., New York: Springer-Verlag

Tytler, D. & Burles, S., 1996, preprint, astro-ph 9606110

Walker, A. G., 1936, *Proc. Lond. Math. Soc.* **42**, 90

Weinberg, D. H., Hernquist, L., Katz, N., & Miralda-Escudé, J., 1996, in *Cold Gas at High Redshift*, ed. Bremer, M., Rottgering, H., Carilli, C., and van der Werf,

P., Kluwer: Dordrecht

Weinberg, S., 1972, *Gravitation an Cosmology: Principles and Applications of the General Theory of Relativity*, John Wiley & Sons

Weymann, R. J., 1993, in *The Environment and Evolution of the Galaxies*, ed. Shull, J. M., Thronson, H. A., Dordrecht: Kluwer

White, S. D. M. & Silk, J., 1979, *Ap.J.* **231**, 1

Zel'dovich, Y. B., 1970, *Astronomy & Astrophysics* **5**, 84

Zel'dovich, Y. B. & Novikov, I. D., 1983, *The Structure and Evolution of the Universe (Relativistic Astrophysics, volume 2)*, University of Chicago Press

Zhang, Y., Anninos, P., & Norman, M. L., 1995, *Ap.J. (Letters)* **453**, L57

Zhang, Y., Anninos, P., & Norman, M. L., 1996, in preparation