NOTE TO USERS

The original manuscript received by UMI contains pages with indistinct print. Pages were microfilmed as received.

This reproduction is the best copy available

UMI
Cosmological Gas-dynamics: Simulations of the Lyman-α Forest

by

James Wright Wadsley

A Thesis submitted in conformity with the requirements for the Degree of Doctor of Philosophy in the University of Toronto

© Copyright by James Wright Wadsley 1998
The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author’s permission.

L’auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L’auteur conserve la propriété du droit d’auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.
Quasar absorption spectra provide the best probe of the state of the gaseous universe at high redshifts. Lyman-α absorption lines map out hydrogen gas structures ranging from tenuous filaments (forest lines) to galaxies (damped absorbers). These are naturally occurring features in hierarchical structure formation theories and well described by the cosmic web picture. A fast non-periodic hydrodynamical N-body code, Tree-P3M-SPH, was developed and extensively tested. Simulations were performed for the standard cold dark matter cosmological model and the hot-cold, open and vacuum extensions. Artificial spectra were generated and fit as observed spectra are, with an automated Voigt profile line fitter, allowing a direct confrontation of theory with data.

Both high resolution and the inclusion of large-scale effects are shown to be essential. These competing aims are satisfied here with many simulations of small patches of the universe and the accurate inclusion of long waves and tides. Mean shear is identified as the critical parameter describing the patches. The results are combined into a representative sample using the theory of Gaussian random fields. The advantages of constrained field initial conditions for high resolution structure formation studies are demonstrated.

The cosmological models studied differ in power spectrum shape, amplitude and dark matter abundances. After rescaling the ultraviolet flux, each model fits the observed flux depression and line column density distribution in detail. Non-local measures, such as two point
functions, would better reflect the highly visible differences between the models. The range of ultra-violet fluxes required falls within the broad observational constraints. Changing the ultra-violet flux history is found to have a measurable impact on the gas temperatures only at high densities. The mini-halo picture of absorbers is confirmed: at redshift $z = 3$, the dominant absorbers are dwarf galaxies for neutral hydrogen columns $N_{HI} \sim 10^{15} - 10^{17}\text{cm}^{-2}$ and filamentary gas below $N_{HI} \sim 10^{15}\text{cm}^{-2}$. Insufficient numerical resolution is shown to suppress or remove the dwarf galaxies. The time evolution, helium absorption, line width versus column density and line width distributions match the observations well for cosmological models with similar fluctuation amplitudes on small scales.
Nor law. nor duty bade me fight.
Nor public men. nor cheering crowds.
A lonely impulse of delight
Drove to this tumult in the clouds.

William Butler Yeats
Acknowledgements

I would like to thank my wife Tara for her patience, assistance and support and my daughter Fiona for mostly sleeping through the night. I would like to thank my supervisor, Dick Bond, for his guidance, insight and openness and my fellow students, members of CITF and friends in Canada. People who have gone out of their way to be friendly and helpful are Omar Lopez-Cruz, James Brown, Pierre Gravel, Hugh Couchman, Felipe Barrientos, Peter Papadopoulos, Marcin Sawicki and Huan Lin. To those I have not explicitly listed, I offer my apologies and my thanks. I gratefully acknowledge the support of a Canadian Commonwealth Scholarship.
Contents

I Introduction ........................................... 1

1 Lyman-α Absorbers ................................. 5

1. Observations ....................................... 5

1.1. Clouds .......................................... 6

1.2. Equivalent Width .............................. 6

1.3. Column Density Distribution ................. 9

1.4. Metal Lines .................................... 16

1.5. Flux Statistics .................................. 19

2. Models of Lyman-α Clouds ...................... 22

2.1. Basic Considerations ......................... 24

2.2. Models ......................................... 25

2.3. Numerical Simulations ....................... 27

References ........................................... 30

2 Method ............................................. 35

1. Introduction ...................................... 35

1.1. Eulerian versus Lagrangian Methods ........ 36
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.</td>
<td>Smoothed Particle Hydrodynamics</td>
<td>38</td>
</tr>
<tr>
<td>2.1</td>
<td>Interpolation</td>
<td>38</td>
</tr>
<tr>
<td>2.2</td>
<td>Derivation of the Equations of Motion</td>
<td>41</td>
</tr>
<tr>
<td>2.3</td>
<td>Variable Smoothing Length</td>
<td>52</td>
</tr>
<tr>
<td>3.</td>
<td>Gas Processes</td>
<td>58</td>
</tr>
<tr>
<td>3.1</td>
<td>Ionization Processes</td>
<td>58</td>
</tr>
<tr>
<td>3.2</td>
<td>Cooling Processes and Temperatures</td>
<td>67</td>
</tr>
<tr>
<td>4.</td>
<td>Gravity</td>
<td>75</td>
</tr>
<tr>
<td>4.1</td>
<td>Particle-Particle Correction</td>
<td>81</td>
</tr>
<tr>
<td>4.2</td>
<td>Fast Fourier Transform for Gravity</td>
<td>82</td>
</tr>
<tr>
<td>4.3</td>
<td>Trees for Particle-Particle</td>
<td>87</td>
</tr>
<tr>
<td>4.4</td>
<td>Force Errors and Method Comparison</td>
<td>90</td>
</tr>
<tr>
<td>4.5</td>
<td>Local Trees for Neighbour Searching</td>
<td>93</td>
</tr>
<tr>
<td>4.6</td>
<td>Particle Setup</td>
<td>97</td>
</tr>
<tr>
<td>4.7</td>
<td>Standard Test Problems</td>
<td>99</td>
</tr>
<tr>
<td>5.</td>
<td>Time steps and Integration</td>
<td>113</td>
</tr>
<tr>
<td>6.</td>
<td>Evolution of the Lyman-α runs</td>
<td>115</td>
</tr>
<tr>
<td>7.</td>
<td>Convergence</td>
<td>116</td>
</tr>
<tr>
<td>Appendices</td>
<td></td>
<td>126</td>
</tr>
<tr>
<td>A.</td>
<td>Estimating the physical analogue of SPH expressions</td>
<td>126</td>
</tr>
<tr>
<td>B.</td>
<td>Anisotropic kernels and Tensor h</td>
<td>127</td>
</tr>
<tr>
<td>B.1</td>
<td>Implementing Tensor H</td>
<td>128</td>
</tr>
</tbody>
</table>
C. Modification to standard SPH Kernels ........................................ 130
D. Density profiles with constant particle mass in 1D ......................... 131
E. Two-Dimensional Gravity ...................................................... 132
References .............................................................................. 134

3 Design ................................................................................. 139
1. Gaussian Random Fields ..................................................... 140
2. Cosmologies ......................................................................... 141
3. Motivation for Constrained Calculations ................................. 144
4. Structure Formation ............................................................. 147
5. Patch statistics ..................................................................... 148
6. Designing Lyman-α Forest Simulations ............................... 152
7. Peak Patches ........................................................................ 156
8. Crafting High Resolution Simulations ................................ 158
9. The Cosmic Web ................................................................... 160
10. Peak Patches at High Redshift .............................................. 164
References .............................................................................. 167

4 Results .................................................................................. 169
1. The Inter-Galactic Medium ................................................. 169
   1.1. Gas Properties ............................................................... 177
2. Spectra ................................................................................. 177
3. Weighting the Simulations ..................................................... 179
4. Flux Statistics and $J_{-21}$ ........................................ 181
4.1. Visual Comparison among Cosmologies .................. 192
5. Voigt Profile Fits .................................................. 192
6. Properties of Gas Associated with Lines ................. 200
7. Column Density Distribution ................................. 200
8. Line Widths ....................................................... 211
9. Rescaling .......................................................... 222
10. Zel'dovich maps ................................................ 226
11. Helium Absorption ............................................ 227
12. A Cloud/Inter-Cloud split .................................. 228

References ................................................................. 238

Conclusion ............................................................... 240
Introduction

Cosmology is a physical science that reaches to the very largest observable scales and to the earliest times in the universe. It takes natural curiosity about origins to the ultimate level. The most popular model of the universe is the big bang (a hot, compact and rapidly expanding beginning). The evidence is substantial: the successful prediction of the light element abundances by big bang nucleosynthesis, the Hubble expansion and the all-pervasive microwave background radiation. The homogeneity of this cosmic background implies that the visible universe must have been causally connected at some earlier time. However, information has not had time to cross the universe during the current mode of expansion. Inflation, a prior period of undecelerated expansion, is invoked to reconcile these facts. The minute anisotropies in the temperature of the background radiation are associated with small density fluctuations in the universe at early times. The fluctuations vary as a function of scale in a way consistent with their being laid down during inflation.

A great success of cosmology has been to show that gravitational instability can take small density fluctuations and, over the age of the universe, build the majestic structure in galaxies and clusters we observe today. Hierarchical clustering is the process by which smaller objects form first (e.g. galaxies) and are then incorporated into larger ones at later times (e.g. clusters). Hierarchical clustering explains the predominance of smaller structures at early times and is a natural consequence of the increasing fluctuation amplitudes we expect as we move to smaller scales.

Observations of the microwave background probe the largest scales. Large scale streaming motions, galaxy correlations, clusters, galaxy groups and individual galaxies probe intermediate scales. Clusters and galaxies are rare events and we see them by the photons they emit. Gravitational dynamics indicate that the density of the universe is significantly higher than the density in ordinary, baryonic matter as determined by big-bang nucleosynthesis. Inventories of baryons indicate very few are in visible objects. Thus, galaxy surveys and cluster observations sample rare, biased baryonic peaks in a sea of dark material. Quasars are ultra-luminous objects
existing up to very high redshifts. Gas clouds seen in absorption in quasar spectra probe the universe in a better way: lines of sight to quasars probe random portions of the inter-galactic medium. We are thus given an unbiased sample of the universe.

The dominant gas in the universe is hydrogen. It causes most of the absorption lines in quasar spectra. The Lyman-α transition of neutral hydrogen at 1216 Å, occurring at high redshifts, is observable at optical wavelengths today. In chapter one we give a broad summary of the observations of Lyman-α absorption in quasar spectra, with extra detail on data that are particularly relevant to this work. We describe early theoretical models based on physical interpretations of the properties of absorption lines and summarize current 3-dimensional numerical studies.

It is difficult to probe density fluctuations at small scales. These fluctuations give rise to dwarf galaxies and the first star clusters, which hierarchical clustering incorporates into galaxies at late times. The numerous, small Lyman-α absorbers are expected to arise on just these small scales. Thus, quasar absorption spectra indirectly probe the first stages of galaxy formation. Metal absorption lines probe the first heavy element production.

Many parameters that describe the cosmology of the universe are poorly constrained. We do not know whether the universe is open or closed. The major constituents are unknown and we are uncertain of the exact baryon density. Our best constraints come from surveys at the current epoch and the microwave background.

Chapter one describes the ever-growing wealth of quasar data available. Good data extends from the current epoch to beyond $z = 4$. We expect the data to contain strong signatures of the underlying cosmology. There has been great progress in our understanding of the high-redshift gaseous universe but the theoretical work to date has yet to provide reliable constraints on the cosmology. The aim of this thesis is to better exploit the Lyman-α absorber data. We achieve this with carefully designed 3-dimensional numerical simulations and detailed comparisons of artificial and real data.
In chapter two we describe the numerical techniques we use to model Lyman-α absorbers. We describe the important physics operating in the inter-galactic medium and justify the assumptions necessary for progress to be made. We give details of the smoothed particle hydrodynamics algorithm and compare several gravity solving methods, including new methods we developed. The final code is named TreeP$^3$M-SPH, which we show by a suite of tests to be accurate and efficient.

Chapter three covers the design of our Lyman-α simulations. We stress the importance of high numerical resolution and the accurate inclusion of large wavelength fluctuations. We detail a method for constructing sets of simulations that meet these criteria. We use four cosmological models that satisfy current constraints: open, vacuum, hot plus cold and standard cold dark matter. The design of the initial conditions for the density perturbation field are fundamental to cosmological numerical work. Our initial condition generation technique derives from analytical descriptions of structure in the universe. Following Bond, Kofman and Pogosyan (1996), we emphasize the importance of peaks in defining the structures present in the universe at all scales. The state of the art in analytical methods is described: the peak patch technique of Bond and Myers (1996). We show how it may be used to model structure on very large scales.

Our strategy is to match our results to the observations by creating artificial spectra and measuring the properties of the lines therein. In chapter four we report simulated results for a range of line properties, including: flux decrement distributions, line column densities in neutral hydrogen, line widths and helium absorption. The results concentrate on the low to moderate column density Lyman-α forest lines. We cannot model high column lines correctly because they require optically thick radiative transfer. In addition, the objects responsible for high column lines are inferred to be strongly star forming, which we do not model. The ultraviolet background radiation ionizes the inter-galactic gas. We modify the ultraviolet levels, the mean baryon density and the numerical resolution and explain the effects on the results. The poorly known ultraviolet background gives us the flexibility to scale our
results to fit the observed data. We show how four cosmological models with similar fluctuation amplitudes at small scales can be made to fit all aspects of the data for low column lines.
Chapter 1

Lyman-α Absorbers

1. Observations

The term “Lyman-α forest”, first put down in print by Weymann (1981), is used to describe the numerous absorption lines blueward of the Lyman-α emission line in quasar spectra. Lynds (1971) discovered the Lyman-α forest and correctly interpreted the lines as due to Lyman-α absorption by neutral hydrogen. Sargent et al. (1980) established the basic properties of the forest. Vast improvements in telescope collecting area and spectrograph sensitivity have produced a wealth of detailed information on absorption lines. In particular, the Keck telescope is producing data of amazing quality (e.g. Kirkman and Tytler (1997), Lu et al. (1996), Kim et al. (1997)), which will soon be matched by the VLT and Gemini telescopes. The current data provide intricate details on metal line systems with implications for the chemical evolution of galaxies (e.g. Songaila & Cowie (1996)), limits on primordial deuterium (e.g. Burles & Tytler (1997), Songaila et al. 1997) and valuable insights into ionizing flux spectrum shape and history through Helium observations (e.g. Hogan et al. (1997)). One important area where good data is starting to become available is real and velocity space correlations (e.g. Williger et al. (1996)).
1.1. Clouds

Lynds (1971) established that most of the absorption features present in the Lyman-\(\alpha\) forest were due to the Lyman-\(\alpha\) transition by correlating with Lyman-\(\beta\). Strictly, the term “forest lines” is used to denote weaker lines. For these lines the column density of neutral hydrogen (the number density integrated along the line of sight) is less than \(N_{HI} \sim 10^{15}\) cm\(^{-2}\) and no associated metal lines are detected. Current surveys with high signal to noise and resolution consistently detect associated metal lines only above \(N_{HI} \sim 10^{14.5}\) cm\(^{-2}\) (Songaila & Cowie (1996)). For some time it was argued whether the lines where all associated with the quasars or intervening cosmological “clouds”. This issue was confused by the large impact the quasar has locally due to its ionizing output, known as the proximity effect, and the existence of associated systems out to several thousand km s\(^{-1}\). Sargent et al. (1980) established the basic features of the line population. It is declining with decreasing redshift to the current epoch, composed of lines fitting an exponential distribution in rest equivalent width and relatively lacking in correlations.

1.2. Equivalent Width

Initial quasar absorption studies with poorer resolution, sensibly focused on equivalent width rather than attempting to fit lines. Equivalent width is the length of continuum that contains the same amount of flux as is absorbed by a line. The rest frame equivalent width follows a curve of growth with gas column as shown in figure 1.1. The growth is linear in \(N_{HI}\) below \(N_{HI} \sim 10^{14}\) cm\(^{-2}\), then as \(\sqrt{\log_e N_{HI}}\) up to \(N_{HI} \sim 10^{18}\) cm\(^{-2}\) and as \(\sqrt{N_{HI}}\) beyond that where the Damping wings of the line profiles are significant. The sensitivity of the line width as a measure of gas column is very low around \(N_{HI} \sim 10^{15}\) cm\(^{-2}\). The equivalent width for a given column scales with redshift \(z\) as \(W(z) = W_{\text{rest frame}}(1 + z)\). Lines are counted as distinct absorption features above some cut-off in rest frame equivalent width. A cut-off of 0.32 Å corresponds to \(10^{14}\) cm\(^{-2}\). There is a small uncertainty in this
Fig. 1.1.— Rest frame equivalent width ($W_0$) curve of growth for Lyman-α absorption lines. The curves represent Doppler line widths: 50, 40, 30 and 20 km s$^{-1}$ from top to bottom, respectively. The dashed curves are the results in the absence of the Lorentz profile convolution.
correspondence because of the effect of thermal broadening. The steep power-law distribution of column densities discussed below dictates that lines near the cut-off dominate the counts.

The population can be roughly fit to the functional form (Murdoch et al. (1986)),

$$\frac{dN}{dz} = A_0 (1 + z)^\gamma$$ and $$\frac{dN}{dW} \propto e^{-W/W^*}.$$  \hspace{1cm} (1.1)

$N$ is the number of lines and $W$ is the rest equivalent width. Typical values at $z = 3$ are $\gamma = 2.3 \pm 0.4$ with $A_0 = 3.5$ (Bajtlik (1992)). Typical values for $W^*$ are $0.22 - 0.35$ Å (Bahcall et al. (1993), Sargent et al. (1980), Weymann (1992)). The region near the quasar must also be excluded to avoid the proximity effect (Bajtlik, Duncan & Ostriker (1988)). Typical exclusion zones are of order 3000 km s$^{-1}$ from the redshift of the quasar. A typical number density is about 100 lines per unit redshift interval at $z \approx 2.5$ above a cut-off of around $W_{\text{min}} = 0.32$ Å. The simple power-law is a poor fit over the whole range $z = 0$ to $z = 4$. Lu et al. (1991) fit two power-laws, with substantially less evolution at lower redshifts. Bahcall et al. (1993) find $\gamma \approx 0.75$ and $A_0 \approx 15$ for $z \lesssim 1$. Most recent HST results give $\gamma \approx 0.85 \pm 0.25$ (Dobrzycki & Bechtold (1997), see figure 1.2). The measured apparent evolution is a function of the continuum fitting procedure and to a lesser degree the minimum equivalent width for the sample. $W_{\text{min}} \approx 0.32$ Å gives $\gamma_{\text{obs}} \sim \gamma_{\text{intrinsic}}$ (Zuo & Bond (1994)). Liu & Jones (1988) performed simulations indicating that with blending and poor resolution, $\gamma_{\text{obs}} < \gamma_{\text{intrinsic}}$ generally. Even with current high resolution ($\sim 0.2$ Å), blending is still severe at higher redshifts ($z \sim 4$). Current studies focus more on line fitting and use extensive simulated data sets to estimate blending effects.

1.3. Column Density Distribution
Fig. 1.2.— Evolution of line number counts with redshift. Note the dramatic increase in counts at around redshift 2.5. A simple interpretation of this figure is a relatively stable population plus a rapidly evolving one that disappears at low redshifts. The data are taken from the references given in the figure.

With sufficient quality data, Voigt profiles may be fit to the absorption features. Line fitting provides both a width, which is interpreted as partly thermal and partly due to velocity broadening, and the neutral column of gas. The physical insight provided by the fitting procedure has been crucial to the development of our understanding of the forest. Most often lines are the result of blends and are not easily fit by a single Voigt profile. The process of selecting the combination of profiles best suited to fit a particular region, often called “deblending”, is not unique. Column density estimations tend to depend on the subjective bias of the observer. For example, a single wide line or several adjacent thinner lines may be used to fit an undulating region in a spectrum. Despite the subjectivity, line fitting remains attractive. Blending is not likely to result in serious systematic errors except at high redshifts and very low columns. Many observers perform fits on simulated
data sets with the same properties as their observed data to understand and control systematic errors \((\text{e.g. } \text{Hu et al. (1995)})\).

The Voigt profile for the Lyman-\(\alpha\) absorption cross-section, \(\sigma\), is given by a convolution of a Doppler profile with the Lorentz or natural line profile.

\[
\sigma(\nu) = C \int_{-\infty}^{\infty} dv \frac{1}{b \sqrt{\pi}} e^{-v^2/b^2} \cdot \frac{\Gamma/(4\pi^2)}{(\nu - \nu_0(1 + v/c))^2 + (\Gamma/(4\pi))^2},
\]

(1.2)

where \(\nu\) is the frequency, \(b\) is the Doppler width, \(\Gamma = 6.25 \times 10^8 \text{s}^{-1}\) and \(C = 3/(8\pi)\Gamma \lambda_\alpha^2\). In this case, \(\nu_0 = c/\lambda_\alpha\) where \(\lambda_\alpha = 1215.6701 \text{\AA}\) and \(c\) is the speed of light.

Below \(N_{HI} \approx 10^{16} \text{cm}^{-2}\), the line profiles are purely Doppler profiles and the convolution with the Lorentz profile has no effect. Above \(N_{HI} \approx 10^{18} \text{cm}^{-2}\), no information about line widths is available because the lines are dominated by the damping wings from the Lorentz part of the profile convolution.

A plot of the observed column density distribution is shown in figure 1.3. Carswell et al. (1984) established a power law relation in the distribution of column densities.

\[
\frac{dN}{dN_{HI}} \propto N_{HI}^{-\beta}.
\]

(1.3)

where \(N_{HI}\) is the column density of neutral hydrogen. This expression fits quite well over the observable range of column densities, \(N_{HI} \approx 10^{12.8} \text{cm}^{-2} - 10^{22} \text{cm}^{-2}\). Values for \(\beta\) range from \(~ 1.7\) (Carswell et al. (1984), Rauch et al. (1992), Lanzetta et al. (1991c)) to more current estimates around \(\beta \approx 1.5\) (Petitjean et al. (1993)). The power-law fit is clearly dependent on the range of columns considered. Recent efforts have focused on correcting for blending, resolution and noise effects with simulated spectra closely matched to high quality observed spectra in order to determine a power law slope for low columns (Hu et al. (1995), Lu et al. (1996)). For lines with columns less than \(N_{HI} \approx 10^{14} \text{cm}^{-2}\), these studies give a remarkably tight value for
Fig. 1.3.— Column density distribution. The plot is derived from the data of Petitjean et al. (1993) and Hu et al. (1995) ($z_{\text{obs}} \sim 2.7$). The boxes represent bins with the top and bottom of each box being the Poisson error. The lower panel contains the same data multiplied by $N_{\text{HI}}^{1.46}$ and renormalized to remove the dominant power-law.
the slope of $\beta = 1.46 \pm 0.06$ (Lu et al. at $z = 3.7$) that does not change down to $z = 2.8$ (Hu et al.). This distribution probably continues down to $N_{HI} \sim 10^{10}$ cm$^{-2}$, providing sufficient numbers of low column systems to explain the observed helium opacity (Kirkman and Tytler (1997)).

As the deviation from a single power is not large, we prefer to plot $\frac{dN}{dN_{HI}}N_{HI}^{-1.46}$ rather than $\frac{dN}{dN_{HI}}$. This removes the several orders of magnitude from the figure to reveal details of the distribution. We have plotted the data this way for the remainder of this work. There is a clear break from the low column power-law at $N_{HI} \sim 10^{14.5}$ cm$^{-2}$. The power-law slope is close to 1.8 at columns above the break. Observationally, this is poorly understood. We will later conclude that it is due to different types of objects producing the absorption above and below this column density.

The fit of Kirkman and Tytler (1997) to the column density distribution has a break in the power law at $N_{HI} \sim 10^{13.5}$ cm$^{-2}$ (see figure 1.4), whereas the larger data set of Petitjean et al. (1993) combined with the data of Hu et al. (1995) for low columns has a break at $N_{HI} \sim 10^{14.5}$ cm$^{-2}$. This result is partially due to the subjective nature of fitting. It is also a resolution effect as the Kirkman and Tytler (1997) data set has resolution (FWHM) of 8 km s$^{-1}$ against $\sim 25$ km s$^{-1}$ for the Petitjean et al. (1993) data. It may also be an intrinsic property of the specific quasar Kirkman & Tytler observed. This possibility is the most worrying because of the large intrinsic variability implied. There is limited published information on intrinsic variability. We generally make our comparisons against the Petitjean et al. (1993) data as it is the more extensive set and has been available longer. The same fundamental features are present in all data sets. The large variations in the mean flux, discussed later, are a strong indicator of changes in the line counts. Cosmic variance and changes in the ionizing flux levels are the likely sources of variations.

One interpretation of the data is that there are two populations, one rapidly dying away at $z \approx 2$ and another slowly evolving to the current epoch (see figure 1.2). For a constant comoving density of absorbers with constant cross-section.
Fig. 1.4.— Column density distribution for the high quality single line-of-sight sample of Kirkman and Tytler (1997). Note that no blending correction has been performed so that the low column slope is not indicative of the underlying distribution.
\[ \frac{dN}{dz} \propto (1 + z)^{1/2} \text{ and } (1 + z)^5 \text{ for a flat Einstein de Sitter universe and an open universe respectively.} \]
The observed low redshift evolution is consistent with either. The evolution of systems of high column density is substantially similar to this at all redshifts. High column systems are strongly associated with large field galaxies. Such a population can not explain the bulk of the high redshift absorption. If we posit a population that is the result of a framework of absorbing structures fixed in comoving space then a given structure will have a number density of neutral hydrogen that evolves as density squared (for photoionization, see Chapter 2) or \((1 + z)^6\) times a length that evolves as \((1 + z)^{-1}\). The neutral column thus evolves as \((1 + z)^5\). With a power-law column density distribution, it follows that the line numbers evolve as \((1 + z)^{5(\beta - 1) - p}\) where \(p = 3/2\) for an Einstein de Sitter universe. For \(\beta = 1.7\) the implied power-law is \((1 + z)^2\); compatible with the observed high redshift evolution. With the addition of some evolution in the intensity of the ionizing flux, our simulations show that this situation well describes the evolving state of the absorbing medium at high redshifts.

A neutral hydrogen column density of \(N_{HI} \gtrsim 10^{17} \text{ cm}^{-2}\) is optically thick to photons shortward of 912 Å, defining Lyman Limit Systems (LLS). LLS systems thus respond differently to the background level of UV flux. The excess of lines at \(N_{HI} \gtrsim 10^{17} \text{ cm}^{-2}\) is likely to be due to this difference. LLS invariably have associated metal lines. It is reasonable to assume that the systems responsible contain large quantities of mostly neutral hydrogen and are star forming. There is of order one LLS per unit redshift interval. For a cross-sectional radius of 10 kpc for Lyman Limit absorption this would imply a comoving number density of \(0.4 \text{ (0.2) } h^{-2}(\text{h}^{-1}\text{Mpc})^{-3}\) at \(z = 2.5\) for a flat matter-dominated (open) universe. This is about an order of magnitude higher than the density of bright galaxies now.

"Damped" Lyman-\(\alpha\) absorbers have column densities \(N_{HI} \gtrsim 10^{18} \text{ cm}^{-2}\) and create sufficiently wide features \((W \approx 7.3(N_{HI}/10^{20} \text{ cm}^{-2})^{1/2} \text{ Å})\) for the radiation damped portion of the Voigt profile to be evident. Lanzetta \textit{et al.} (1991c) demonstrate a substantial excess, by a factor of 10, of damped systems relative to
a single power-law fit extrapolated from $N_{HI} < 10^{18} \text{cm}^{-2}$. From the data presented in figure 1.3, however, it could be argued that there is a deficit of systems at intermediate columns.

The columns measured in the range $10^{16} \text{cm}^{-2} < N_{HI} < 10^{19} \text{cm}^{-2}$ are probably the most uncertain and there are small numbers of objects in the high column bins. Lanzetta (1991b) performed extensive modelling of systems in this column range with the CLOUDY photo-ionization code of Ferland et al. (1992), with results indicating that a smooth power-law in the underlying distribution of total gas columns could easily give rise to the irregular observed distribution of neutral hydrogen columns.

There is less information available on Lyman Limit Systems than other column ranges. Storrie-Lombardi et al. (1994) determine an exponent $\gamma = 1.55 \pm 0.45$ for the evolution (see equation 1.1) from a relatively large sample. Their data also allows two power-law evolution: $\gamma_{z<3} \approx 1.0$ and $\gamma_{z>2.5} \approx 2.8$. Lanzetta (1991b) puts $\gamma_{z>2.5} \approx 5.7$. Bahcall et al. (1993) derive $\gamma \sim 0.5$ from HST data for $z \lesssim 1$. Thus the data is consistent with no evolution at low redshift. This probably indicates a large contribution from a second population at higher redshifts.

Lanzetta et al. (1993,1991) acquired a large dataset of damped Lyman-α systems, using IUE data. The data indicates $\gamma = 1.15 \pm 0.55$. Whilst the number of systems is not strongly evolving, the column density in individual absorbers declines with decreasing redshift. Given that field spirals never exceed $N_{HI} \approx 10^{21} \text{cm}^{-2}$ (van Gorkom (1992)) and that damped systems are associated with primordial spiral galaxies, it is not surprising that one does not find systems with $N_{HI} > 10^{21} \text{cm}^{-2}$ for $z < 1.5$. Bahcall et al. (1993) found only one damped system in their 36 quasar spectra at $z \lesssim 1$. Lanzetta et al. (1995) calculated the cosmological density of neutral gas in absorbers, finding a near exponential decline since $z = 3$, when the mass in neutral gas was comparable to the current stellar population: $\Omega_\star \approx 2.9 \pm 1.2 \times 10^{-3} h^{-1}$. 
The line widths are characterized using the Doppler parameter \( b = (2kT/m_H) \). Bulk motions in the cloud can increase \( b \) beyond that resulting from the gas temperature alone. The distribution of Doppler parameters for Lyman-\( \alpha \) clouds fit a Gaussian profile with \( b_{\text{mean}} \approx 30 \text{ km s}^{-1} \) and \( \sigma_b \approx 10 \text{ km s}^{-1} \) (Carswell et al. (1991)). The corresponding temperatures are \( T \approx 2 \times 10^4 \text{ K} \) up to \( 1 \times 10^5 \text{ K} \). A typical value is taken as \( T = 3 \times 10^4 \text{ K} \). Exceedingly low Doppler parameter measurements (\( \lesssim 10 \text{ km s}^{-1} \) implying temperatures \( < 10000 \text{ K} \)) were reported for the Lyman-\( \alpha \) forest (Pettini et al. (1990)) but are probably metal lines or noise. Kirkman and Tytler (1997) find that their very high quality data are consistent with no low temperature lines. Voigt profiles for high column densities provide no Doppler parameter information due to the dominance of radiative damping, though associated metal lines are often available in these cases.

In figure 1.5 we show Keck high resolution line width data from Kim et al. (1997) and Hu et al. (1995). Hu et al. are also the source of our low end column density distribution data. There are 91 and 132 lines total in the low and high Kim et al. sample respectively and 170 and 285 lines total in the high and low column bins in the Hu et al. sample respectively. The difference between the two column density bins is slight, and not particularly significant with a sample this small. We expect some difference because lower density, cooler gas is associated with lower columns.

1.4. Metal Lines

The metal lines associated with systems with \( N_{HI} \gtrsim 10^{15} \text{ cm}^{-2} \) have shed light on their metal abundances and structure. Typical values for the metallicity of these systems are \( \sim 0.01 \text{ Solar} \) with enhanced Si/C by around a factor of 3 as is observed in population II stars (Songaila & Cowie (1996)). Metal lines are not reliably detected in systems of lower column density, and thus it is unclear whether low column density systems contain primordial cosmological abundances. A species
Fig. 1.5.—Hu et al. (1995) line width data for two column density cuts, with $<z> = 2.8$. Kim et al. (1997) line width data within the column density range. $N_{HI} = 10^{12.8} \, \text{cm}^{-2} - 10^{14.3} \, \text{cm}^{-2}$ at two redshifts. Both sets of observations were done with the Keck telescope.
which should be quite detectable at the temperatures and densities expected for low
column systems is OVI (1032 Å). It is difficult to detect because it lies within the
Lyman-α forest. The number of thin lines seen in the data is consistent with the
expected number from noise with no strong indication of a contribution from OVI
lines. Thus a very low level of metallicity at low columns is indicated (Hellsten et al.
(1997)).

Carswell et al. (1991) determined metallicities of 1/10 solar in careful study of
a damped system (\(N_{HI} \sim 10^{18} \text{ cm}^{-2}\) ) at \(z \approx 2\), which is in line with other estimates
(Foltz et al. (1988), Lanzetta (1992)). Carswell et al. found that the metal systems
broke up into several lines, with a total spread of \(\sim 150 \text{ km s}^{-1}\). They consider the
individual Doppler parameters to be consistent with bulk velocities rather than
thermal broadening (See also Lanzetta & Bowen (1992), Wolfe et al. (1993)). Recent
data show that a range of metallicities is present in these systems, with 1/10 solar
being typical.

MgII line systems represent a similar population to LLS (Sargent et al. (1988b)).
CIV line systems, however, have a larger cross-section than LLS (assuming they are
produced by the same population) and evolve more actively, which may be related
to increases in metallicity (Sargent et al. (1988a)). Wolfe et al. (1992) examined a
\(z = 2.466\) damped system in a narrow band centred on Ly-α and estimated a size
of \(24 \text{ h}^{-1}\text{kpc} \times 11 \text{ h}^{-1}\text{kpc}\). The physical sizes of Ly-α forest clouds are difficult
to constrain, but multiple quasar studies have determined that \(R_{\text{cloud}} \sim 150 \text{ h}^{-1}\text{kpc}\nwith a range of \(60 - 500 \text{ h}^{-1}\text{kpc}\) from Monte Carlo studies. Filamentary objects
with cross section of order \(100 \text{ h}^{-1}\text{kpc}\) with typical lengths \(\geq 1 \text{ h}^{-1}\text{Mpc}\) fit the data
extremely well (Fang et al. (1996)).

Steidel (1992) surveyed galaxies near quasar lines of sight and found a galaxy for
every HI line with associated MgII absorption for \(0.3 \leq z \leq 0.9\). He measured impact
parameters and found a mean of \(23 \text{ h}^{-1}\text{kpc}\) with no systems beyond \(40 \text{ h}^{-1}\text{kpc}\). This
corresponds nicely to estimates of the MgII radius of galaxies required to explain
the observed line numbers (see also Lanzetta (1992)). Morris et al. (1993) searched
along the line-of-sight to 3C 273 for galaxies associated with Ly-α lines, with the conclusion that the lines could not all be connected to "normal" luminous galaxies but that a substantial galaxy-absorber correlation existed, though much less than the galaxy-galaxy correlation. Metal line systems are known to correlate similarly to galaxies (Steidel (1992)) whereas measures of correlation for the Ly-α forest observe a mild excess at $\Delta v \sim 150 \text{ km s}^{-1}$ (Webb (1987). Zuo & Bond (1994)). Meiksin & Bouchet (1995) found correlations on scales $100 - 600 \text{ km s}^{-1}$ with a maximum amplitude of 0.5-1.

1.5. Flux Statistics

Treating the Lyman-α absorbers as discrete clouds is only likely to be accurate for a population of isolated compact objects. For the Lyman-α forest in particular, the gas in which the clouds arise is quite likely to be only moderately condensed relative to the mean density and to possess structure in velocity space. This is especially true of the tenuous gas that contributes the Helium opacity. The flux decrement is defined.

$$D_A = \langle 1 - F \rangle. \tag{1.4}$$

where $F$ is the normalised flux transmission (Observed flux as a fraction of the extrapolated continuum level). The flux decrement avoids the subjectivity of line fitting and provides an overall measure of the absorption. It is measured between 1216 Å and 1026 Å in the rest frame, the region between Lyman-α and Lyman-β. Thus it is a measure of absorption due to Lyman-α with a small degree of contamination by metal lines. The contribution to $D_A$ is dominated by lines around $N_{HI} \sim 10^{14} \text{ cm}^{-2}$. This connection is made explicitly in figure 1.6, using the data of Kirkman and Tytler (1997). This spectrum has a mean flux decrement of around $D_A = 0.31$ at $\langle z \rangle = 2.7$.

The power-laws are rough fits to the column density distribution shown in
Fig. 1.6.— The contribution of lines of different column to the mean opacity due to Lyman-α absorption. The data is from single line-of-sight sample of Kirkman and Tytler (1997)
figure 1.4. The diamonds with the rough curve running through them are the actual contributions to the opacity from lines in the spectrum. The curve labelled $N^{-1.46}$ is included to demonstrate that a single power-law cannot reproduce the result. The curves are normalized to have the same area underneath each one.

The flux decrement measurements are very sensitive to the accuracy of the continuum fit. There also appears to be large intrinsic variability among lines of sight. Recent estimates for the flux depression are given in table 1, with data taken from Rauch et al. (1997b), Press et al. (1993), Schneider et al. (1991) and Zuo & Lu (1993).

<table>
<thead>
<tr>
<th>Z</th>
<th>Rauch</th>
<th>Press</th>
<th>Schneider</th>
<th>Zuo</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>&lt; 0.05</td>
<td>&lt; 0.05</td>
<td>&lt; 0.02</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.148</td>
<td>0.15</td>
<td>0.09-0.16*</td>
<td>0.2</td>
</tr>
<tr>
<td>3</td>
<td>0.316</td>
<td>0.36</td>
<td>0.25-0.38*</td>
<td>0.35</td>
</tr>
<tr>
<td>4</td>
<td>0.543</td>
<td>0.62</td>
<td>0.50-0.64*</td>
<td>0.55</td>
</tr>
</tbody>
</table>

The ranges given are rough estimates covering the range of values derived by the authors for a number of different quasars at the given redshift. All the authors exclude at least the region within 10,000 km s$^{-1}$ (rest frame) of the quasar redshift. Schneider et al. (1991) presented a summary of early work with low resolution spectra and no attempt to remove contamination by metal lines. The figures for this entry are given to the nearest 0.05. Press et al. (1993) performed detailed continuum fitting based on each spectrum redward of the Lyman-α emission. Inspection by eye of the continuum fits indicates a high probability of substantial errors in the extrapolation to lower wavelengths, especially considering the crude functional form employed: $C_1/2 \lambda^{1/2} + C_1 \lambda$. They do not present their data in such a way that a range of values may be determined directly. The ranges in the table, denoted with an asterisk, are based on the error bars quoted for their fit. Zuo & Lu (1993) did not measure their $D_A$ from spectra but from published line lists. They excluded metal systems. They tested this method versus mean flux measure from spectra.
and concluded that low column absorbers that do not appear in line lists are not important. This assumption is probably much less valid at lower redshifts which may partially explain the relatively low values they get. Rauch et al. (1997b) use recent Keck data. The continuum is well estimated using regions of the spectrum apparently free of absorption. Finding such regions becomes problematic at \( z \sim 4 \).

Only 7 quasars comprise their total sample compared to compared to \( \sim 10 \) per redshift bin in the other studies. Given the large variance indicated in the larger data sets, the accurate Rauch et al. results must unfortunately be considered indicative rather than conclusive. Reasonable estimates of the values at \( z = 2, 3 \) and 4 are probably \( D_A = 0.15, 0.3 \) and 0.6 respectively. As noted in the table there is not much low redshift data available – particularly around \( z = 1 \). A measure closely related to \( D_A \) is the one point distribution of fluxes. This distribution is shown for 3 redshifts in figure 1.7. The data in Kim et al. (1997) show that the intrinsic variation between two entire lines of sight at redshift 2.3 may be large, around 0.1 on the figure. The variability is expected to be much less by reshift 3. This reflects the behaviour of the flux depression. The distributions extend outside the range 0 to 1, because of the noise in the normalized spectra.

These measures change in a predictable way when the ionizing flux level is modified. They provide a way to make predictions about the flux level from simulations. In particular, constraints have been estimated for the combination \( (\Omega_b h^3/J_{-21}) \) which reflects the neutral hydrogen density per unit redshift and thus the mean opacity for Lyman-\( \alpha \) absorption. Here, \( \Omega_b \) is the critical density in baryons and \( J_{-21} \) is ionizing flux in units of \( 10^{-21} \text{erg cm}^{-2} \text{s}^{-1} \text{Hz}^{-1} \text{Sr}^{-1} \) measured at 912 Å.

This is strictly true only for pure photo-ionization. The uncertainty in the data and the ultra-violet flux levels make a strong limit on \( \Omega_b \) difficult to obtain.

2. Models of Lyman-\( \alpha \) Clouds

2.1. Basic Considerations
Fig. 1.7.— Flux level distribution. The data shown are from Rauch et al. (1997b). Observations of 7 quasars went into this sample. The intrinsic variation between entire lines of sight is large at redshift 2.3, at least 0.1 on the figure.
The parameter space for Lyman-α cloud models is fairly well constrained. Current estimates of the UV flux at high redshifts put $J_{-21} \sim 1$ and definitely $\gtrsim 0.01$ (Bechtold (1992)). This flux is consistent with that expected from observed quasars. The total ultra-violet flux contribution from galaxies is poorly constrained but is probably dominant at very high redshifts, $z \gtrsim 5$. This flux enforces a high degree of ionization:

$$\frac{n_{HI}}{n_H} \approx 0.1 \left( \frac{T}{10000 \text{ K}} \right)^{-0.7} n_H (\text{cm}^{-3}) J_{-21}^{-1}. \quad (1.5)$$

for the number density of hydrogen $n_H << 1$. The cosmological mean H density is around $10^{-5} \text{cm}^{-3}$ at $z = 3$. As previously noted, Doppler parameter measurements put $T_{\text{cloud}} \sim 3 \times 10^4 \text{K}$. A thermal and a Gaussian distributed bulk velocity component would add Doppler parameters such that $b_{\text{obs}}^2 = b_{\text{bulk}}^2 + b_{\text{thermal}}^2$, indicating that the thermal component of the Doppler parameter could easily be $\sim 10 \text{ km s}^{-1}$ lower than observed (ie. $T_{\text{cloud}} \sim 2 \times 10^4 \text{K}$). These values are entirely consistent with photo-ionization as the dominant source of heating. Other sources of heating, such as supernova energy, shock heating and softer ionizing radiation from early stellar populations, are less well constrained.

The Gunn-Peterson test for HI limits the amounts of neutral hydrogen present in the IGM by measuring the minimum absorption. A typical limit for the optical depth due to diffuse gas is $\tau_{\text{GP}} \lesssim 0.02$ at $z \approx 4$ for HI not associated with clouds. (Giallongo et al. (1994). $\tau_{\text{GP}} \lesssim 0.04$ in Webb et al. (1992)). This constrains the amount of HI in the diffuse IGM: $n(HI) \lesssim 4.5 \times 10^{-12} (\tau_{\text{GP}}(z)/0.02) \Omega^{1/2} h((1 + z)/5)^{3/2} \text{cm}^{-3}$ (Peebles (1993)). For smoothly distributed gas this limits the total $n_H \lesssim 5 \times 10^{-6} (T/10^4 \text{K})^{-0.35} J_{-21}^{-0.5} \text{cm}^{-3}$ at $z \approx 4$. The mean density of $n_H = 1.8 \times 10^{-5} (\Omega_b h^2/(0.0125))((1 + z)/5)^3 \text{cm}^{-3}$.

Cold dark matter clumps without regard to pressure and can be expect to be gravitationally dominant in all but extremely dissipated systems. The gas pressure
inhibits gas structure formation. The Jeans scale at the mean density is given by.

\[ r_J \approx 0.14 \left( \frac{4}{1 + z} \right)^{1/2} \left( \frac{T}{3 \times 10^4 \text{ K}} \right)^{1/2} \Omega_{nr}^{-1/2} h^{-1} \text{ comoving Mpc} \]  

where \( \Omega_{nr} \) refers to the fraction of the critical density in non-relativistic matter at the current epoch. \( h \) parameterizes the influence of the Hubble constant \( H_0 = 100h \text{ km s}^{-1} \text{ Mpc}^{-1} \). The Jeans mass is of order \( 3 \times 10^9 (4/(1 + z))^{3/2} h^{-1} \Omega_{nr}^{-1/2} \text{ M}_\odot \). The linear expression can only be a rough guide, especially in the presence of deep dark matter potential wells. An alternate estimate of the effects of gas pressure is based on velocities. The typical ion velocities in the gas are around \( 27 \sqrt{T/30000} \text{ km s}^{-1} \), inhibiting gas collapse in halos with less than this velocity dispersion. The gas temperatures do not vary sufficiently strongly as a function of density for cooling instabilities to be expected to play an important role in the general medium.

An important consideration is the intrinsic cloud-cloud and galaxy-cloud correlations in the forest. If the clouds are associated with galaxies or smaller mass objects in a hierarchical model, a measurable correlation is expected. Models in which the clouds are simply a denser phase of the IGM expect no correlation. The data is fairly conclusive about the strong clustering of Lyman-limit absorbers with galaxies at low redshift.

2.2. Models

The first attempts to model the Lyman-\( \alpha \) forest invoked the idea of a hot, confining IGM (Sargent et al. (1980), Ostriker & Ikeuchi (1983), Ikeuchi & Ostriker (1986)). The clouds are typically 100 times denser than the IGM and have \( T_{\text{cloud}} \approx 3 \times 10^4 \text{ K} \). The huge range of cloud masses required to model the entire range of column densities precludes a single value for the IGM pressure. This model has lost prominence with the success of cosmological structure formation models in which pressure is not the dominant mechanism. The ideas behind this model can be
redirected to clouds in the hot halos of galaxies. High column clouds are observed in the Galaxy as "high velocity" clouds. At some level, the outer parts of galaxies must make a contribution to the Lyman-\(\alpha\) forest absorption. Pressure confined clouds within the hot gas of galaxy halos have been estimated by Mo (1994) to contribute roughly \(1/5\)th of the Ly-\(\alpha\) forest.

The low column clouds are quite likely to be unbound. Free expansion has been invoked to explain possible low \((b < 10\ \text{km}\ \text{s}^{-1})\) Doppler parameter measurements (Duncan \textit{et al}. (1991)), but both the effectiveness of an adiabatic cooling mechanism is doubtful and later observations have tended to rule out significant numbers of lines arising in cold gas (Rauch \textit{et al}. (1993), Rauch \textit{et al}. (1992)).

Dark matter minihalos were the first attempt to tie Lyman-\(\alpha\) clouds into the hierachical structure formation framework (Rees (1986), Bond \textit{et al}. (1988)). It is a basic prediction of the CDM and other hierachical cosmological models that there should be a large number of sub-galactic mass objects collapsing at early redshifts. These objects are expected to be subsumed into larger objects at lower redshifts.

Mini-halo models predict that the gas in dark matter halos behaves differently at different radii. Gas in the cores \((N_{\text{H}I} \gtrsim 10^{15}\ \text{cm}^{-2})\) will be in hydrostatic equilibrium. This gas may reach high densities and substantially recombine in the larger objects \((N_{\text{H}I} \gtrsim 10^{19}\ \text{cm}^{-2})\), being shielded from the UV flux. The size of this region is highly susceptible to any variation in the UV flux (Murakami \& Ikeuchi (1993), Murakami \& Ikeuchi (1994)). The rest of the cloud consists of a dynamic region not in thermal equilibrium and an accretion layer joining onto the IGM (Meiksin (1994)). In general there will be a radius at which there is a cross-over between accretion and outflowing winds, dependent on the depth of the dark matter well. (Meiksin (1994), Charlton \textit{et al}. (1994)). These "photo-ionization winds" only begin once the IGM is photo-ionized (Bond \textit{et al}. (1988)). They apply for fairly small halos. Bond \textit{et al}. and Meiksin employed one-dimensional hydrodynamics codes. Bond \textit{et al}. (1988) noted that collapsed Lyman-\(\alpha\) cloud scale objects \((\sim 100\ \text{kpc})\) are abundant \((2h^3\ \text{Mpc}^{-3})\) at \(z \gtrsim 3.5\) and predict a fairly abrupt decline in the number.
of clouds by $z \approx 2$ if photo-ionization becomes important at that epoch. Current data indicate substantial photo-ionization had occurred as early as $z \sim 5$ so that very small collapsed masses are probably not an important source of clouds. Recent attempts to re-invent the idea (Abel & Mo (1997)) run into the same problem.

There has been a fundamental shift in our understanding of structure formation with the application of three-dimensional hydrodynamical $N$-body codes to the problem. This has been combined with analytical progress on the prevalence of non-spherical modes of collapse in the medium. The fundamental addition to the picture is filamentary networks as the dominant contributor to Lyman-α forest absorption.

2.3. Numerical Simulations

Cen et al. (1994) was the first work that focused very high resolution three-dimensional codes specifically onto the question of the Lyman-α forest. It was performed with the Eulerian code of Cen (1992). Several techniques have now been applied to the problem.

Table 2 summarizes the main Lyman-α codes and runs in print in 1997. Mass Resolution refers to the spacing of the initial condition grid for the dark matter. Where a (2) is placed alongside, the listed resolution will be degraded by roughly a factor of 2 by the smoothing inherent in the PM method used to evolve the simulation. "Volume" is the usable high resolution volume of the simulation. Gas and (Dark) Resolution refer to the best physical resolution at $z = 2$ in gas and dark matter respectively. $\lambda_{\text{MAX}}$ refers to the largest wavelength modelled with dark matter particles in the simulation. Periodic initial conditions have /2 alongside to indicate the vast reduction in available waves for an FFT based initial condition. For example, in practice, periodic simulations with a strong fundamental mode are not used. Shown in brackets are waves that are applied but evolved without particles directly, e.g. with a linearly evolved tidal field.
The key features lacking in many numerical approaches are a good sample and high resolution. Single periodic box simulations cannot provide a representative sample of the universe. They truncate large scale power and additionally rely on one realisation to create their simulated sample. Notably in the work of Zhang et al. (1996), their high resolution region was void-like. Some workers use larger boxes at the expense of resolution (Hernquist et al. (1996), Mücket et al. (1996)). The importance of gas physics and small scale structures make such simulations less relevant for the Lyman-\(\alpha\) forest. The best compromise between size and resolution among the periodic simulations listed is that detailed in Miralda-Escudé et al. (1996) and uses the Cen TVD code. Steinmetz has applied his free boundary code to explore absorbing systems in a realistic environment with the focus on metal lines (Rauch et al. (1997a)).

All the numerical codes have managed to reproduce the gross features of the Lyman-\(\alpha\) forest. This situation probably owes more to the fact that gas with a correlation function of \(\sim r^{-2}\) readily produces the column density distribution observed, as was noted by Rees (1986). However, with most simulations the line width distributions and more subtle features of the forest absorbers are also reproduced. These results are sensitive to the physics of the heating of the medium and the accuracy of the simulated hydrodynamics. We discuss the important results to date in Chapter Four.

In the face of this challenging numerical problem, we elected to perform multiple high resolution simulations with constrained initial conditions that cover the parameter space of possible patches of the universe. With smaller simulations that run in a reasonable time, we were able to run several cases exploring different cosmologies and variations in other parameters such as the ultra-violet flux history. The details of the code and the physics we include are discussed in Chapter Two. The design of the simulations is covered in Chapter Three.
Table 2: Simulator Groups and Lyman-α runs. See text for details.

<table>
<thead>
<tr>
<th>Simulator(s)</th>
<th>Mass Resolution (Mpc)</th>
<th>Volume (Mpc$^3$)</th>
<th>Gas Volume (kpc)</th>
<th>(Dark)$\lambda_{MAX}$ (Mpc)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wadsley &amp; Bond</td>
<td>0.1</td>
<td>65 (x5)</td>
<td>1 (1)</td>
<td>12.8 (6000)</td>
</tr>
<tr>
<td></td>
<td>0.2</td>
<td>65 (x5)</td>
<td>1 (1)</td>
<td>12.8 (6000)</td>
</tr>
<tr>
<td></td>
<td>0.31</td>
<td>2000 (x5)</td>
<td>3 (3)</td>
<td>40 (6000)</td>
</tr>
<tr>
<td>SCGM; HCDM, ΛCDM, OCDM; $\Omega_b = 0.0125h^{-2}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50$^3$ + 40$^3$ + 32$^3$ gas, same dark particles. TreeP$^3$M-SPH free boundary with external tides (this work)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Steinmetz</td>
<td>0.11</td>
<td>6 (x4)</td>
<td>1.25 (2.5)</td>
<td>30/2</td>
</tr>
<tr>
<td></td>
<td>0.19</td>
<td>27 (x4)</td>
<td>2.5 (5)</td>
<td>30/2</td>
</tr>
<tr>
<td>SCGM; $\Omega_b = 0.0125h^{-2}$; $\sigma_8 = 0.7$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20$^3$ gas, 20$^3$ dark + tidal particles. Grape-SPH free boundary with external tides, Rauch et al. (1997a)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cen Ostriker</td>
<td>0.032</td>
<td>98</td>
<td>6 (11)</td>
<td>4.6/2</td>
</tr>
<tr>
<td></td>
<td>0.064</td>
<td>98</td>
<td>11 (21)</td>
<td>4.6/2</td>
</tr>
<tr>
<td></td>
<td>0.11</td>
<td>3600</td>
<td>18 (35)</td>
<td>15.4/2</td>
</tr>
<tr>
<td>ΛCDM; $\Omega_b = 0.015h^{-2}$; $\Omega_{DM} = 0.4$. $\Omega_{\Lambda} = 0.6$; $\sigma_8 = 0.8$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>288$^3$ gas/PM cells, 144$^3$ dark particles for first case, TVD PM-Eulerian Miralda-Escudé et al. (1996)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hernquist Katz</td>
<td>0.35</td>
<td>11000</td>
<td>20 (20)</td>
<td>22.22/2</td>
</tr>
<tr>
<td>SCDM; $\Omega_b = 0.0125h^{-2}$; $\sigma_8 = 0.7$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>64$^3$ gas, 64$^3$ dark particles, Periodic Tree-SPH Davé et al. (1996)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mucket Kates</td>
<td>0.1</td>
<td>2100</td>
<td>50 (50)</td>
<td>128/2</td>
</tr>
<tr>
<td>SCDM; No gas; $\sigma_8 = 0.7$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>128$^3$ dark particles, 256$^3$ PM code Mück et al. (1996)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zhang Norman</td>
<td>0.075(2)</td>
<td>880</td>
<td>25 (25)</td>
<td>9.6/2</td>
</tr>
<tr>
<td></td>
<td>0.019(2)</td>
<td>14</td>
<td>6 (6)</td>
<td>9.6/2</td>
</tr>
<tr>
<td></td>
<td>0.050(2)</td>
<td>33</td>
<td>8 (8)</td>
<td>3.2/2</td>
</tr>
<tr>
<td></td>
<td>0.013(2)</td>
<td>0.5</td>
<td>2 (2)</td>
<td>3.2/2</td>
</tr>
<tr>
<td>SCDM; $\Omega_b = 0.015h^{-2}$; $\sigma_8 = 0.7$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>128$^3$ gas/PM cells, 128$^3$ dark particles, Hercules(Zeus) PM-Eulerian Zhang et al. (1996)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


REFERENCES


Burles. S. & Tytler, 1997, accepted to Astr. J.


Ferland. G. J. 1989, *Ohio State University Internal Report* 89-001


Lin, H., Kirshner, B., Shectman, A., Landy, S., Oemler A., Tucker. D. and
Miralda-Escudé, J., Rauch, M., Sargent, W., Barlow, T., Weinberg, D., Hernquist,
IAP Colloquium: Structure and Evolution of the IGM from QSO Absorption
Line Systems, eds. P. Petitjean, S. Charlot


Chapter 2

Method

1. Introduction

In this chapter we discuss the numerical solution of both gas physical processes and gravity. Our hydrodynamical method, smoothed particle hydrodynamics (SPH), is described in detail, with discussion of our implementation. We then detail our treatment of ionization, heating and cooling processes. We provide a full discussion of our gravity solvers and some comparisons between schemes. Details particular to the design of our calculations for the Lyman-α forest are presented in Chapter Three.

For this thesis we performed a series of calculations with an evolving SPH-Nbody code. The original code was equipped with a fairly competitive implementation of the SPH algorithm but a poor gravity solver. Particle-Mesh. Much work over the course of the Ph.D. has gone into redeveloping the code, so that the simulations we performed were not only competitive with those other workers, but were good enough to give us accurate answers in a reasonable time. The original code did not meet these criteria. The major coding effort went into the gravity code, resulting first in P³MG, a full multigrid gravity solver with a particle-particle correction for high resolution. The method proved limited in speed and force accuracy, which
prompted a switch to a fast Fourier transform for the particle-mesh part of the force. Particle-particle forces proved to be the final hurdle, necessitating a tree-code solution for the particle-particle force. The current code is a full order of magnitude faster than the first particle-particle high resolution versions. This has allowed us to perform a large number of simulations for many cosmologies.

1.1. Eulerian versus Lagrangian Methods

It is often argued that Eulerian simulations are superior for studies of the IGM, particularly the Lyman-\(\alpha\) forest. Eulerian simulations have a grid with fixed comoving resolution for the gas phase. They rely upon particles for dark matter, with a particle-mesh scheme for the gravity. This means that the dark matter mass resolution is never better than 2 grid cells – the typical value quoted for particle-mesh (Hockney and Eastwood (1988)). This means the dominating mass, and thus the gravitational potentials are determined with a method with constant mass resolution and poor spatial resolution. (In dense regions, the poor gravity resolution for the mesh compromises the method.) The usual comment is that these peaks are unimportant for problems such as the Lyman-\(\alpha\) forest. We find this to be untrue, with peaks providing an important contribution (Chapter Four). The alternate claim is that voids are simulated particularly well. This only true for the gas – the dark matter resolution in the voids degrades as they empty. We find that all structures in the IGM have an important dark matter component and thus the additional gas resolution in the voids is of no consequence.

Eulerian simulations possess a distinct advantage of speed per cell. Lagrangian techniques typically require significantly more computation per resolution element (particle or cell). Typical Eulerian simulations are performed with 4 times as many cells per dimension as smoothed particle simulations (SPH), \(i.e.\) \((4n)^3\) cells versus \(n^3\) particles. Thus at mean density the gas resolution for Eulerian codes is typically of order 4 times better. Eulerian techniques can more easily implement high order
gas solvers, such as the Piecewise Parabolic Method. PPM (Woodward and Collela (1984)). This allows better resolution with of order 2 cells required to resolve shocks versus 4 particle spacings for SPH (Monaghan and Gingold (1983)). Thus up to overdensities of order a few tens (or even a few hundred for complex high order schemes) an Eulerian code with \((4n)^3\) cells can resolve gas as well or better than SPH with \(n^3\) particles for similar computation. For dark matter the difference is much less. The PM methods relied upon by grid calculations smooth structure below 2 grid cells. whereas \(P^3M\) or Tree based codes typically follow the mass as it clusters down to order of \(1/20^{th}\) of the initial spacing. This implies that pure Lagrangian codes are superior for overdensities of \(\sim 10\) or better. It is just these overdensities that create the dominant observable structures in QSO absorption spectra (see Chapter Four). The deficiencies of Eulerian schemes can be substantially addressed through adaptive schemes. These schemes adaptively improve the resolution in gas and gravity – most important for cosmological problems. The usual criterion for local adaptation to improve resolution is to maintain constant mass resolution. Such schemes are extremely complex, generally not publicly available and required several years to implement. They have been shown to be effective for simulating clusters (Bryan & Norman 1997).

A key concern for cosmological hydrodynamics is the extent to which it is important to match the resolution in the dark matter and the gas. Our SPH scheme matches resolution in these components by its construction. Eulerian schemes rely on fairly distinct techniques for gas and dark matter. We argue that additional gas resolution is wasted if the underlying dark matter structures are not resolved. For this not to be the case, we would have to see substantial separation between the species – large variation in the ratio of gas to dark matter. Such separation only occurs in strongly dissipative situations, such as the formation of galactic disks and is not a factor in Lyman-\(\alpha\) forest simulations. Another attractive feature of gas particle simulations is superior local conservation of angular momentum. Grids tend to diffuse it rapidly. This is not important outside highly collapsed regions such as
2. Smoothed Particle Hydrodynamics

Smoothed Particle Hydrodynamics (SPH) is an approach to hydrodynamical modelling devised by Lucy (1977). It is a particle method that does not refer to grids for the calculation of hydrodynamical quantities: All forces and fluid properties are found through interpolation over nearby particles. By employing particles that move with the fluid, SPH removes the need for advective terms in the equations of motion and is thus a fully Lagrangian method. These factors make SPH comparatively non-diffusive and flexible in its handling of irregular geometries. The majority of the early development of SPH was carried out by Gingold and Monaghan. Monaghan further developed the method, applying it to novel situations such as incompressible fluids and relativistic fluid motion (Monaghan (1992)). The development of SPH for use in cosmology required the development of variable smoothing to handle huge dynamic ranges and fast gravity solvers (e.g. Hernquist and Katz (1989)).

The original SPH concept was simple. SPH was a method which was computationally fast and also relatively easy to program. It still retains these features, though many improvements have been made and are detailed in the following sections. For example, it is straightforward to add complex equations of state, heating and cooling terms, abundance variation and other physics to SPH. The chief modification to SPH over its early form has been variable smoothing lengths. Without this change the Lagrangian nature of SPH is compromised because the spatial resolution, determined by the smoothing length would remain fixed.

2.1. Interpolation

The basis of the SPH method is the representation and evolution of smoothly varying quantities whose value is only known at discrete points in space. This is
similar to finite differences except that the points do not need be and in general will not be regularly ordered. This characteristic led to SPH being described as a Monte Carlo type method with the large errors $O(1/\sqrt{N})$ associated with that method (Gingold and Monaghan (1977)). It has been shown by Monaghan (1985a) that the method is more closely related to interpolation theory with errors $O((\ln N)^d/N)$. where $d$ is the number of dimensions.

The advantage of a fully Lagrangian scheme is the absence of diffusion associated with calculating advective terms using grids. Steinmetz and Müller (1993) find that this advantage only becomes apparent in more than one dimension and that otherwise SPH possesses more intrinsic diffusion than one-dimensional finite difference Lagrangian schemes.

The integral form for the interpolation procedure is as follows:

$$\langle f(\mathbf{r}) \rangle = \int W(\mathbf{r} - \mathbf{r}', h) f(\mathbf{r}') d\mathbf{r}'$$

(2.1)

where $f$ represents a physical variable such as density or internal energy. $W(\mathbf{r} - \mathbf{r}', h)$ is known as the kernel function and it effectively performs the interpolation. $h$ is the smoothing length scale for the interpolation. $W$ should satisfy the conditions:

$$\int W(\mathbf{r}, h) d\mathbf{r} = 1$$

(2.2)

and

$$\lim_{h\to 0} W(\mathbf{r}, h) = \delta^{(3)}(\mathbf{r}).$$

(2.3)

This second condition ensures that the method goes to the correct continuum limit.

In general $W$ is chosen to be spherically symmetric and positive such that:

$$\langle f(\mathbf{r}) \rangle = \int W\left(\frac{|\mathbf{r} - \mathbf{r}'|}{h}\right) f(\mathbf{r}') d\mathbf{r}'.$$

(2.4)
which approximates the the value of \( f \) at \( r \) as follows:

\[
\langle f(r) \rangle = f(r) + c(\nabla^2 f)h^2 + O(h^3),
\]

(2.5)

where \( c \) is independent of \( h \) (Benz (1990)).

There exist additional conditions which help to optimize the role of the kernel. These are discussed in section 2.2.

The conversion to a summation form is straightforward:

\[
f_s(x_i, h) = \sum_{j=1}^{n} \frac{f_j}{n_j} W(r_i - r_j, h),
\]

(2.6)

where \( f_j \) is the discretised value of \( f \) at \( r_j \) and \( n_j \) is the particle number density at \( r_j \). \( 1/n_j \) in the sum plays the role of \( dr \) in the integral form. The errors involved in this conversion can be estimated by the Poisson summation formula. For kernels which are basis splines, labelled \( M_n \) (Schoenberg (1973)), the errors are \( O(h^n) \) (Monaghan (1985a)). Gingold and Monaghan (1977) show that the expectation value of the sum is the integral form.

Normally particles are assigned masses \( m_j \) and \( m_j/\rho_j \) is written in place of \( 1/n_j \), where \( \rho_j \) is the density at \( r_j \). There is nothing wrong with the concept of individual masses for each particle. Simulations with constant mass particles and particles with varying masses tend to give identical results. The quality of the interpolation procedure becomes poor if the masses vary sharply.

Attempting to maintain a uniform number density is a good aim for fixed smoothing length codes. Monaghan and Varnas (1988) experimented with a method where if a single cell on a regular grid laid over the particles held to many or too few particles, then all the particles in that cell were replaced by a set number of new particles. Mass, energy and all other properties of the old particles were interpolated to the new particles. This approach re-introduces the interpolation related diffusion of grid based methods. In some cases, particles could also be assigned very small
or very large masses. The authors did not encounter any major problems with this approach.

It has been found in situations with large amounts of mixing, that heavy particles can move among the lighter particles and crash through them like bowling balls (Herant (1993)). One beneficial aspect of varying mass is having a uniform number density of particles and thus resolution. In the abovementioned situation, however, a single heavy particle may move to a low density region while many light particles take its place in a high density region which completely defeats the purpose. In our simulations we employ a range of masses so that we can model the effects of large scale tides without great computational expense. We therefore take great care to ensure that the larger mass particles are excluded when we calculate our simulation results (see Chapter Four). Resolution limitations with fixed $h$ lead naturally to the idea of $h(r)$, a smoothing length that varies in space. The errors introduced by varying $h$ are not easy to estimate or compensate for and will be discussed in section 2.3.

There are two interpretations of formula 2.6. Each particle may be considered to be scattering its data over space with a distribution given by $W$ or it may be assumed that data is being gathered from surrounding particles with weightings given by $W$ (Hernquist and Katz (1989)). Of course, this distinction only becomes important if $h$ is not constant but varies in space as $h(r)$. There is no consensus on the most appropriate interpretation. We prefer the latter interpretation since the value of a variable at the particle is unambiguously the value of the continuous version of the variable at that point in space. Unfortunately with variable smoothing length a hybrid of the two schemes must be used to ensure momentum conservation (see section 2.3.).

2.2. Derivation of the Equations of Motion

*The Continuity Equation*
The SPH equations are discrete versions of the standard hydrodynamical equations. It can be shown that the SPH equations represent these equations with errors that are $O(h^2)$, which is consistent with the accuracy of the interpolation method. These derivations generally assume constant $h$. It can be shown to be generally true for operators such as $\partial/\partial t$, $\nabla$ and $\mathbf{v} \cdot \nabla$ that the following relations are correct to $O(h^2)$ if $h$ is constant and $f$ is smoothly varying:

$$\langle X(f(r)) \rangle = X(\langle f(r) \rangle) = X(f(r)), \tag{2.7}$$

where $X$ represents an operator such as $\partial/\partial t$, $\nabla$ or $\mathbf{v} \cdot \nabla$. For example, the Continuity Equation:

$$\frac{d\rho}{dt} = -\rho \nabla \cdot \mathbf{v}. \tag{2.8}$$

when convolved with the kernel in the smoothed interpolation procedure becomes:

$$\frac{d\langle \rho \rangle}{dt} = -\nabla \cdot (\langle \rho \mathbf{v} \rangle) + \langle \mathbf{v} \rangle \cdot \nabla \langle \rho \rangle. \tag{2.9}$$

Translating the integrals to sums, the SPH form for the Continuity Equation is derived:

$$\frac{d\rho_i}{dt} = \sum_{j=1}^{n} m_j (\mathbf{v}_i - \mathbf{v}_j) \cdot \nabla_i W_{ij}. \tag{2.10}$$

Here $W_{ij}$ is shorthand for $W(r_i - r_j, h)$. Equation 2.8 was rearranged slightly so that the final form would be Galilean invariant. Rearrangement of the equations of motion in order to produce a better SPH form is a common procedure in SPH. Some of the most robust expressions in common usage, such as the form for thermal conduction, are the result of intuition and experimentation rather that direct derivation.
If $dm_i/dt = 0$ for all $i$, then:

$$\rho_i = \sum_{j=1}^{n} m_j W_{ij},$$  \hspace{1cm} (2.11)$$

is an exact solution of 2.10. Thus, density may be calculated using 2.11 and exact mass conservation will be ensured. This equation is unable to model sharp boundaries as density will gradually drop off to zero at the boundary over a length scale of $\sim 2h$. We do not use information within several smoothing lengths of our simulation boundaries and thus avoid such effects.

**The Momentum Equation and Pressure forces**

The momentum conservation equation without dissipation,

$$\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla)\mathbf{v} = -\frac{\nabla P}{\rho},$$  \hspace{1cm} (2.12)$$

when convolved with the kernel, $W$, becomes for constant $h$:

$$\frac{d\mathbf{v}}{dt} = -\int \frac{\nabla P}{\rho} W(|\mathbf{r} - \mathbf{r}'|, h) d\mathbf{r}' - \oint \frac{P}{\rho} W(|\mathbf{r} - \mathbf{r}'|, h) dS',$$  \hspace{1cm} (2.13)$$

where $\oint$ denotes a surface integral. The surface term disappears if $W$ or $P$ vanishes on the boundaries, which is acceptable for most astrophysical problems. Boundaries for which this is not the case are commonly dealt with using ghost particles. Ghost particles are particles temporarily created near boundaries to ensure periodicity or otherwise prevent problems. If the pressure beyond the boundary is known, it can be subtracted from the pressure of the particles before the summation is performed to correctly model the boundaries.

The $-\nabla P/\rho$ term has many representations in SPH. To conserve linear momentum, this term must be symmetric between all particle pairs, i.e. symmetric
in $i$ and $j$. This is generally true of expressions for $dv_i/dt$. The most common form is:

$$\frac{dv_i}{dt} = -\sum_{j=1}^{n} m_j \left( \frac{P_i}{\rho_i^2} + \frac{P_j}{\rho_j^2} \right) \nabla_i W_{ij},$$  

(2.14)

where $\nabla_i$ denotes the gradient with respect to $x_i$. Other forms exist, but this is the most robust (Monaghan (1992), Benz (1990)).

There are alternate ways of convolving the equations of motion with the smoothing kernel. One is to derive the equations from an action principle. Gingold and Monaghan (1982) write down a Lagrangian for SPH:

$$L = \sum_{j=1}^{n} m_j \left( \frac{1}{2} \mathbf{v}_j^2 - U(\rho_j) \right),$$  

(2.15)

where the internal energy $U = U(\rho(\mathbf{r}))$ is a function of density only. They state that this approach ensures accurate energy and exact momentum conservation even if $h(\mathbf{r})$ is spatially varying.

**The Energy Equation**

A form derived from equation 2.14 can also be used to represent $-(P/\rho)\nabla \cdot \mathbf{v}$ in the Energy equation:

$$\frac{du}{dt} = -\frac{P}{\rho} \nabla \cdot \mathbf{v},$$  

(2.16)

producing the SPH form:

$$\frac{du_i}{dt} = \frac{1}{2} \sum_{j=1}^{n} m_j \left( \frac{P_i}{\rho_i^2} + \frac{P_j}{\rho_j^2} \right) (\mathbf{v}_i - \mathbf{v}_j) \cdot \nabla_i W_{ij}.$$  

(2.17)

The combination of 2.14 and 2.17 conserves total energy exactly. Given an expression for $dv_i/dt$, it is straightforward to gain the related expression for $du_i/dt$ that results in total energy conservation. Exact energy conservation can also be understood
in terms of energy exchange between particles. From the preceding formulas it is straightforward to show that the energy change for particle $i$ due to particle $j$ is given by:

$$\frac{dE_{\text{tot},i}}{dt} \big|_j = -m_im_j A_{ij} \nabla_i W_{ij} \frac{1}{2} (v_i + v_j).$$

(2.18)

$A_{ij}$ is symmetric in $i$ and $j$ and contains the pressure, density and viscosity information. $\nabla_i W_{ij}$ is anti-symmetric in $i$ and $j$ and thus $dE_{\text{tot},i}/dt \big|_j = -dE_{\text{tot},j}/dt \big|_i$.

Numerical experiments have indicated that 2.17 can lead to negative internal energies, $u_i$. This led Hernquist and Katz (1989) to use a form based on $\sqrt{P_i/P_j}/\rho_i\rho_j$ for both equations. This did not, however, remove the problem completely. The problem with 2.17 is that $P_j/\rho_j^2$ at neighbouring particles contributes to the change in energy at particle $i$. Negative energy can result due to particle $i$ having a substantial negative $du_i/dt$ term even when $P_i/\rho_i^2$ and thus $u_i$ at the particle is close to zero.

Benz (1990) first demonstrated that the form:

$$\frac{du_i}{dt} = \frac{P_i}{\rho_i^2} \sum_{j=1}^n m_j (v_i - v_j) \cdot \nabla_i W_{ij},$$

(2.19)

when combined with 2.14, still resulted in exact conservation of energy, but without the problem of negative internal energies. We make use of this formulation in our calculation, as it is simple and effective.

The nature of the SPH equations is such that it easy to include more physics by simply adding terms. Problems can arise, however, with cooling and heating terms. In many astrophysical problems, the cooling timescale is often much shorter than the hydrodynamical timescale. In these cases, a prohibitively small time step can be avoided through implicit integration of the energy equation, provided sufficient care is taken in design so that accuracy is maintained. Most applications use variations of the Newton-Raphson algorithm for this purpose (Hernquist and Katz 1989).
Monaghan and Varnas (1988)). We employ an implicit second order integrator for the energy $E(t)$:

$$\frac{E(t+\Delta t)-E(t)}{\Delta t} = PdV + (\text{Heat} - \text{Cool}) (T(E^*), \rho).$$

$$E^* = \frac{E(t)+E(t+\Delta t)}{2}. \quad (2.20)$$

The predicted midpoint values for $\rho$ and the $PdV$ work are used.

The continuum energy equation including thermal conduction is:

$$\frac{du}{dt} = -\frac{P}{\rho} \nabla \cdot v + \frac{1}{\rho} \nabla \cdot (\kappa \nabla u). \quad (2.21)$$

Thermal diffusion presents a difficult problem due to the necessity of approximating a second derivative. A robust expression for $1/\rho \nabla \cdot (\kappa \nabla u)$ (Brookshaw (1986)) is:

$$-\sum_{j=1}^{n} \frac{(q_i + q_j)(u_i - u_j)}{\bar{\rho}_{ij}(r_{ij}^2 + \eta^2)} \mathbf{r}_{ij} \cdot \nabla_i W_{ij}, \quad (2.22)$$

where $q = \kappa/\rho$, $r_{ij} = r_i - r_j$, $\bar{\rho}_{ij} = \frac{1}{2}(\rho_i + \rho_j)$ and $\eta^2 \approx 0.01h^2$ acts to prevent singularities. In practice it is possible to avoid the need for $\eta$ in the actual implementation because $r_{ij} \cdot \nabla_i W_{ij}/r_{ij}^2 = 1/r \partial W_{ij}/\partial r$. The right hand side can be tabulated and for the $M_4$ spline kernel, does not possess a singularity at the origin. This expression is useful in the modelling of shocks, treated in section 2.2.

**Viscosity and the treatment of shocks**

The earliest motivation for the inclusion of viscosity within the SPH formalism was to simulate shocks. SPH particles behave to some extent as fluid molecules do. Real fluid molecules are heated as a result of the passage of shocks, gaining high random velocities. This is manifested on the macroscopic scale by an increase in internal energy. Early SPH simulations of shocks resulted in large post-shock oscillations. High mach number shocks also resulted in large amounts of particle
interpenetration. Monaghan and Gingold (1983) introduced artificial viscosity to SPH. It transfers energy present in unphysical macroscopic fluid motion into internal energy. Finite difference codes have approached this problem with artificial viscosities based on $\nabla \cdot \mathbf{v}$. The idea was to increase the effective pressure to $\sim \rho v^2$ at the shocks instead of $\rho c^2$ in the vicinity of the shock. Artificial viscosity based on SPH expressions for $\nabla \cdot \mathbf{v}$ do not perform well in practice because they are not sufficiently local measures of fluid motion (Hernquist and Katz (1989)). The following term is commonly used for the purpose:

$$
\Pi_{ij} = \begin{cases} 
-\frac{\alpha \delta_{ij} \mu_{ij} + \delta \mu_{ij}^2}{\rho_{ij}} & \text{for } \mathbf{v}_{ij} \cdot \mathbf{r}_{ij} < 0, \\
0 & \text{otherwise.}
\end{cases}
$$

$$
\mu_{ij} = \frac{h(\mathbf{v}_{ij} \cdot \mathbf{r}_{ij})}{r_{ij}^2 + \eta^2}.
$$

(2.23)

where $\alpha$ and $\beta$ are coefficients of terms representing a linear viscosity and a Von Neumann-Richtmyer viscosity, respectively. $\eta^2 \approx 0.01h^2$ is included to prevent singularities. Using only a linear term in $\mu$ has been found to be insufficient for strong shocks. The optimal value for these coefficients varies somewhat, depending on what is being simulated. For a Gaussian kernel, $\alpha \approx 1$ and $\beta \approx 2$ are standard choices and are what we employ (Monaghan (1992)). The $\Pi_{ij}$ term is included in the equations of motion as follows:

$$
\frac{d\mathbf{v}_i}{dt} = -\sum_{j=1}^{n} m_j \left( \frac{P_i}{\rho_i^2} + \frac{P_j}{\rho_j^2} + \Pi_{ij} \right) \nabla_i W_{ij},
$$

$$
\frac{d\mathbf{u}_i}{dt} = \frac{1}{2} \sum_{j=1}^{n} m_j \left( \frac{P_i}{\rho_i^2} + \frac{P_j}{\rho_j^2} + \Pi_{ij} \right) (\mathbf{v}_i - \mathbf{v}_j) \cdot \nabla_i W_{ij}.
$$

(2.24)

An alternate formulation (Monaghan (1985a)) is:

$$
\frac{d\mathbf{v}_i}{dt} = -\sum_{j=1}^{n} m_j \left( \frac{P_i}{\rho_i^2} + \frac{P_j}{\rho_j^2} \right) (1 - \alpha \mu_{ij} + 3 \mu_{ij}^2) \nabla_i W_{ij}.
$$

(2.25)

Both forms perform similarly. They conserve linear and angular momentum and are zero for solid body rotation. There is, however, an unwanted shear component to the
viscosity (see appendix A and Monaghan (1992)). Equation 2.25 performs better in high mach number simulations. The optimal values of $\alpha$ and $\beta$ here also vary from $\sim 1$ for low mach number ($\leq 5$), to higher values for high mach number (Lattanzio et al. (1986)).

Simulators have had much difficulty producing realistic galactic discs (Navarro and White (1994)). Part of the problem is that the unwanted shear component to the viscosity transports angular momentum incorrectly. Balsara et al. (1989) proposed a modification to 2.23 designed to remove the effects of artificial viscosity in pure shear flows, by introducing a switch $f_i$:

$$f_i = \frac{|\nabla \cdot \mathbf{v}|}{|\nabla \cdot \mathbf{v}| + |\nabla \times \mathbf{v}|}$$  \hspace{1cm} (2.26)

but with suitable adjustments to avoid singularities. This term is zero for a pure shear flow where $\nabla \cdot \mathbf{v} = 0$, and one for irrotational flows. This form is of particular importance for the modelling of discs generally. Our numerical experiments indicate that it is difficult to create a switch that does not have at least slight adverse affects on the handling of shocks (see section 4.7.). As disc modelling is not our aim and shocks are quite important in the inter-galactic medium, we decided not to include any switch.

Another problem resulting from the use of artificial viscosities is wall heating. Wall heating is a slightly incorrect distribution of thermal energies near shocks, due to the different treatment of thermal and kinetic energy by SPH. The thermal conduction equation 2.22, can be used to ameliorate this effect, with $q_i + q_j$ replaced by $h_g(\bar{\epsilon}_{ij} + 4 |\mu_{ij}| )$ for $\mathbf{v}_{ij} \cdot \mathbf{r}_{ij} < 0$ (Monaghan and Lattanzio (1991)). They suggest values for $g$ ranging from $\sim .25$ for weaker shocks to $\sim .5$ for high mach number ($\geq 10$) calculations.

Artificial viscosities have not been completely successful in reducing inter-particle penetration and tend to smear shocks out over several particles. Monaghan (1988) also notes that artificial viscosities substantially alter the effective Reynolds
number.

Godunov's method has been successful is overcoming the need for artificial viscosity in grid based methods (Godunov (1959)). The method involves treating boundaries between cells as discontinuities and solving the Riemann problem exactly at this point. A higher-order version known as the Piecewise Parabolic Method is generally considered to be the state of the art in finite difference hydrodynamical modelling (Woodward and Collela (1984)). Attempts to implement the method in SPH have led to problems associated with momentum conservation. Inutsuka (1993) has developed an approach which introduces ghost particles along the direction of $\nabla P$, before and after the true particle, and solves the one-dimensional Riemann problem. He has modelled mach 170 spherical shocks in three dimensions with this method. The method is unwieldy and is not expected to perform well in problems with less symmetry.

Monaghan (1989) sought to reduce penetration of particles within shocks without resorting to large artificial viscosities. To this end he introduced a correction term to the motion equation, $\frac{dr_i}{dt} = v_i$ which brings the particle motion closer to the local average velocity:

$$\frac{dr_i}{dt} = v_i + \epsilon \sum_{j=1}^{n} \frac{m_j}{\rho_{ij}} (v_j - v_i) W_{ij},$$

with $0 \leq \epsilon \leq 1$. Using $\epsilon = 1$ prevented any sort of penetration. He calls this approach XSPH. This correction term does not affect angular or linear momentum or energy conservation and is Galilean invariant. Monaghan did find, however, that it reduced the group velocity for information travelling at longer wavelengths.

Balsara et al. (1989) approach the XSPH modification as a spatial filter for velocity. They suggest a small value $\sim 0.1$ for $\epsilon$. Values of $\epsilon$ near unity have been found to adversely affect the results of test problems examined in Wadsley (1993). We do not make use of XSPH.
A key feature of SPH is the interpolation performed by the kernel. The kernel is a smooth function of separation that determines the spatial distribution of each particle’s data. Schoenberg (1973) places interpolation into two categories: Ordinary interpolation for which \( f_s(x_i, j) = f_i \) and smoothing interpolation for which \( f_s(x_i, j) \neq f_i \). Smoothing interpolation is designed to deal with noisy data and is ideal for particle methods. This is discussed in Monaghan (1985a), an excellent source of analysis regarding interpolation in particle methods. It is the smoothing aspect which lends SPH its name. Interpolation techniques are fundamental to this work. We rely on the robust interpolating procedures employed by the SPH method to transfer all our particle information to grid lines of sight to produce artificial spectra for analysis.

Aside from the necessary conditions of normalisation, equation 2.2, and a correct continuum limit, equation 2.3, there many other considerations which go into the design of a good kernel.

For example, there is often a value for the smoothing length, \( h \), of the order of the interparticle separation, which optimizes the accuracy of the interpolation. Also, beyond one or two particle separations it is desirable that the kernel should go to zero or become negligible. This minimizes the number of particle-particle interactions that must be calculated. Once a sufficient number of nearby particles are included for a good sample of the local values, extra particles add to the amount of computation and also compromise the local character of the interpolated value. The exponential kernel, an early choice in the literature, little used now suffers from this problem:

\[
W(u) = \frac{1}{8\pi h^3} e^{-u}, \quad \text{with} \quad u = \frac{|r|}{h}.
\]  

(2.28)

An important consideration for good interpolation is that the kernel vary
smoothly. In this way errors due to particle disorder are minimized. Thus, the first and second derivatives of the kernel should be continuous. A continuous second derivative ensures that changes in the inter-particle force due to small changes in the particles’ positions are small.

A robust choice of kernel is the Gaussian kernel of Gingold and Monaghan (1977):

$$W(u) = \frac{1}{(\sqrt{\pi}h)^d}e^{-u^2}. \quad (2.29)$$

where $d$ is the number of dimensions. This kernel suffers a little from its extended nature. It is extremely useful for analysis because it is easy to take derivatives. It interpolates well with very small errors if $h$ is larger than the interparticle separation. From a computational standpoint the disadvantage of this kernel is that it drops to a negligible level beyond $3h$, effectively require a summation over 100-200 particles even for fairly uniform particle distributions.

Probably the best choice of kernel for most applications is the $M_4$ spline kernel of Schoenberg (1973), introduced to SPH by Monaghan (1985a):

$$W(u) = \frac{\sigma}{h^d} \begin{cases} 1 - \frac{3}{2}u^2 + \frac{3}{4}u^3, & 0 \leq u \leq 1, \\ \frac{1}{4}(2 - u)^3, & 1 \leq u \leq 2, \\ 0, & u \geq 2, \end{cases} \quad (2.30)$$

where the normalisation constant $\sigma$ takes the values:

$$\frac{2}{3}, \quad \frac{10}{\pi}, \quad \text{and} \quad \frac{1}{\pi} \quad (2.31)$$

in one, two and three dimensions respectively. It has continuous first and second derivatives and compact support so that only particles within $2h$ of the summation point contribute to the interpolated sum. It interpolates exactly for linear functions in one dimension if $h$ is the interparticle separation and as very small errors otherwise. These splines have also found use in gravity solvers, but usually with
lower order splines.

The Fourier transforms of the Schoenberg splines are monotonically decreasing and smooth. This fact makes them highly useful for smoothing initial conditions because they are compact but unlike the top-hat, they have minimal ringing in $k$-space. The expressions for the Fourier transform of the $M_4$ spline in one and three dimensions are.

\[
\begin{align*}
\bar{W}_{1D}(k) &= h \left[ \frac{\sin(\pi k h)}{\pi k h} \right]^4, \\
\bar{W}_{3D}(k) &= \frac{12}{(k h)^5} (-2\sin(k h) + \sin(2k h)) + k^{-1}[6 - 8\cos(k h) + 2\cos(2k h)].
\end{align*}
\]

We discuss some modifications of this kernel intended to address specific minor shortcomings of the method in appendix C. We use the standard $M_4$ kernel in all our calculations.

### 2.3. Variable Smoothing Length

Spatial variation of the smoothing length, $h(r)$, is an essential modification to the SPH technique for simultaneous high numerical resolution on varied length scales with only a limited number of particles. Fixed $h$ sets upper and lower limits on density. We lose gas properties entirely if the interparticle spacing becoming significantly larger than $h$.

The major problem with introducing variable $h = h(r, t)$ is the new terms introduced into the SPH equations. There are two approaches to this. Most often the correction terms are assumed to be small and thus neglected (with fair justification, Evrard (1988)). Alternately, one can attempt to correct for these terms. There is clearly a large error introduced into the gravitational potential if $h(r, t)$ is used for the gravitational softening. During a cosmological simulation, the values of
If $h_{\text{grav}} = h(\mathbf{r}, t)$ this implies a lessening of the potential energy and energy non-conservation. For this reason, we follow the majority of simulators and use a fixed $h_{\text{grav}}$ (e.g. Evrard (1988), Hernquist and Katz (1989)). Corrections for the variation have been included with some success (Hernquist (1993), Nelson and Papaloizou (1993)). Different smoothing lengths for the gravity and hydrodynamics could conceivably lead to unwanted effects. We maintain a well defined potential energy by fixing our gravitational smoothing at a physical value of $h_{\text{grav}}$. Steinmetz and Müller (1993) also indicate that variable $h$ results in increased numerical diffusion relative to fixed $h$, though still smaller than finite difference schemes in multi-dimensional calculations.

Most current implementations of SPH use isotropic smoothing for each particle: the alternative is to allow for ellipsoidal kernels with a tensor describing the smoothing length as a function of direction. Such a method is difficult to implement satisfactorily (see appendix B).

To implement variable $h(\mathbf{r}, t)$, each particle retains its own smoothing length $h_i$. Momentum conservation requires that the force between particles be symmetric, and thus all terms including $h$ in the SPH summations must be symmetrised with respect to the two particles. One common approach is to use $h = \bar{h}_{ij} = \frac{1}{2}(h_i + h_j)$ in the equations (e.g. Benz (1988), Evrard (1988)) so that:

$$W_{ij} = W(\mathbf{r}_{ij}, \frac{1}{2}(h_i + h_j)).$$

(2.34)

Alternately, the kernel may be symmetrised by taking an average (Hernquist and Katz (1989)):

$$W_{ij} = \frac{1}{2} (W(\mathbf{r}_{ij}, h_i) + W(\mathbf{r}_{ij}, h_j)).$$

(2.35)

Monaghan (1992) stated that the latter form follows from an action principle formulation. Hernquist and Katz point out that this form is an average of the scatter and gather interpretations of the interpolation procedure. They write down
correction terms but do not include them in their code. Benz (1990) and our own experiments indicate no strong difference in the results from the two methods. The particle distribution to represent the same density distribution will be slightly different. We find that the kernel averaging method tends to require more neighbours for good SPH summation estimates of variables and thus more computation. An advantage to averaging the $h$ values is that the range of influence of each particle is smoothly defined as a function of position, if the $h$ are smooth. Conceptually this formulation can be understood in terms of a mapping from an initially uniform particle distribution. The value of $h$ then reflects the determinant of the Jacobian of the coordinate transformation to the current state. This presents a simple way to set up an initial $h$, density distribution for one dimension, for which the Jacobian is trivial.

The dense clumps expected in cosmological simulations tend to result in rapid changes to $h$, in space and often in time as clumps move through the medium. Larger numbers of neighbouring particles help maintain stable values of $h$ but need more computation. The original suggestion was 32 neighbours. the number of neighbours on a regular grid with a spacing of $h$, but simulators increasingly favour larger values, mainly to ensure than with variation the number is not too low. Often particles will have large numbers of neighbours contributing practically nothing and no nearby neighbours. This is a pathological case of highly clumped simulation for any method of $h$ estimation that relies on counting neighbours. An excellent solution is to weight the neighbours in a manner that reflects their contribution and thus count the contributions—not the neighbours. A first guess might be to use the kernel itself for the weighting, except that it too centrally peaked and counts a few very near neighbours as sufficient. The problem for neighbour counting is large numbers of particles near $2h$, the maximum interaction distance. We achieved best
results with the function.

\[
W(u) = \frac{1}{3.65833\pi h^3} \begin{cases} 
  1, & 0 \leq u \leq 1. \\
  1 - \frac{1}{6}(u - 1)^2 + \frac{1}{3}(u - 1)^3, & 1 \leq u \leq 1.5. \\
  (3 - u)^3, & 1.5 \leq u \leq 2. \\
  0, & u \geq 2.
\end{cases}
\]  

(2.36)

This function is uniform out to \( h \), where particle make strong contributions to direct summed estimates and gradient estimates, and tails off where the contribution is low. We selected our ideal value for the sum of the contributions, \( \Sigma \), to be that resulting from 50 neighbours for a uniform distribution, \( \Sigma_{50} \). We calculate this sum and iteratively change \( h \) for each particle at the end of each step according to.

\[
h_{\text{new}} = h_{\text{old}}(\Sigma_{50}/\Sigma)^{1/3}. 
\]  

(2.37)

This scheme must be iterative because the solution is an implicit function of all \( h \)'s simultaneously: The range of influence of a particle in any direction is half determined by the \( h \)'s of the particles in that direction. We also use this method to generate \( h \)'s for an SPH estimate of the dark matter density distribution at every major output. It takes only \( \sim 6 \) iterations for the \( h \) estimates to converge to within a few percent of the final answer, starting from a uniform estimate. Our average number of neighbours is less than 50, due to the particle distribution. As long as few particles have less than 30 we can be confident that the interpolation is done well.

If errors due to neglecting correction terms for \( h(r, t) \) are to be kept small, then \( h(r, t) \) must not change rapidly on small scales (\( \sim h \)). Similarly, \( h(r, t) \) should not change rapidly in time. Any scheme that sets or updates \( h \_i \) should take this into account. Codes which vary \( h \) in time and space are known to have slightly larger errors in conservation laws (Hernquist (1993)).

The time step criteria (section 5.) restricts the fractional rate of change of the particle velocities and positions each step so as to maintain accuracy during the integration. With a scaling analysis, the criteria imply that the smoothing length
for a given particle should never change by more that $\sim 10\%$. There are always a few particles in any given step (representing less that 0.1% of the total) that will approach clumps and desire large changes due to shortcomings of the algorithm. since it only makes one correction per step. We prevent particle smoothing lengths from changing more than 10% in one step. In the absence of this limit these particles' $h$'s tend to oscillate in value. The rms change in a given step is around 2% initially which is entirely due to keeping up with the fast expansion between steps at high reshift and quickly settles down to $< 1\%$ at $z < 10$.

Alternate $h$ setting schemes

The optimal value of $h(r, t)$ is proportional to the average local interparticle separation and thus related to density:

$$\frac{h(r, t)}{h_0} = \left( \frac{\rho(r, t)}{\rho_0} \right)^{\frac{1}{d}}. \quad (2.38)$$

where $d$ is the number of dimensions and $h_0$, $\rho_0$ are constants. This leads naturally to a partial differential equation for $h(r, t)$ which may be integrated with the other equations of motion (Benz (1990)):

$$\frac{dh}{dt} = -\frac{1}{d} \frac{h \, d\rho}{\rho \, dt} = \frac{1}{d} h \nabla \cdot \mathbf{v} \quad \text{from the continuity equation.} \quad (2.39)$$

This scheme provides for a smooth variation in $h(r, t)$ in time and space. Near shocks density changes rapidly but with errors proportional to the gradients in $h$ we want to avoid sharp changes in $h$. We have found that adding a small amount of artificial diffusion produces improved results. The full equation for $h_i$ is then:

$$\frac{dh_i}{dt} = -\frac{1}{d} \frac{h_i}{\rho_i} \sum_{j=1}^{n} m_j \mathbf{v}_{ij} \cdot \nabla_i W_{ij}$$
The diffusion term is similar to the form for artificial thermal diffusion 2.22. \( \kappa \) should be \( \sim 1 \). We include the shock related \( \mu_{ij} \) term because sharp changes in \( h_i \) occur near shocks and need to be smoothed. The continuum limit of the diffusion term is proportional to \( \nabla^2 h \).

This scheme works well for low density contrasts in fairly uniform density environments. For cosmology, this scheme tends to fail because errors in the gradient terms systematically underestimate the rate of change in \( h \) required. Some workers perform periodic corrections to \( h \) to maintain a good number of neighbours. This scheme relies upon an estimate of the divergence which is not a quantity SPH estimates to high accuracy.

In an investigation of the effect of varying \( h \), Hernquist (1993) calculated the necessary correction terms for energy conservation for \( h = h(t) \). He notes that highly accurate energy, entropy and mass conservation do not seem simultaneously possible. He also notes that integrating the energy equation leads to large errors in entropy and vice versa. He does not distinguish errors in entropy as being better than errors in energy conservation. He advocated integrating the entropy equation, so that these unavoidable errors are visible as non-conservation of energy. With the heating and cooling present in our calculations, our simulations are in a near equilibrium temperature regime and we integrate energy as it is more appropriate.

Nelson and Papaloizou (1993) attacked the problem of errors due to variable \( h \) with correction terms by treating the SPH particles as a Hamiltonian system. They use the following definition:

\[
\begin{align*}
  h_i &= G \left( \sum_{j=1}^{n} \chi_{ij} H(| \mathbf{r}_i - \mathbf{r}_j |) \right), \\
  &= \frac{1}{2} | \mathbf{r}_i - \mathbf{r}_{\text{furthest neighbour}} |, 
\end{align*}
\]  

(2.41)
in the framework of a Hamiltonian system. \( G \) and \( H \) are functions and \( \chi_{ij} \) are coefficients. The authors derived a force term including correction terms for a variable \( h \) updated with the iterative neighbour counting method of Hernquist and Katz (1989). The first expression is a general form and the second expression, resulting from a simple choice of the functions \( G \) and \( H \) was used in their code. The authors reported a reduction in the errors in energy conservation by a factor of 5 for spherical collapse simulations. These corrections proved computationally expensive for the added accuracy.

3. Gas Processes

3.1. Ionization Processes

The ionization rates in the IGM are such that it is in ionization equilibrium for all but the lowest density gas. The lowest density gas (well below mean density, in voids) may fall out of equilibrium but does not contribute to the Lyman-\( \alpha \) forest hydrogen absorption lines and thus has no impact on those results. We discuss the importance of this for Helium absorption in Chapter Four. We use equilibrium abundances in all our calculations. This has several advantages – it eliminates the necessity to store the additional information on the ionized species between steps and it provides stable answers. The short timescales for ionization processes in reasonably high density regions combined with the high degree of ionization means that great care must be taken with non-equilibrium codes. We can also change the photo-ionization level very easily when we create artificial spectra or other data for comparison with observation. We use this to investigate the appropriateness of modifying the ultraviolet flux or density from the self-consistent value used during a calculation. This is covered in detail in Chapter Four.

The dominant ionization processes in the inter-galactic medium are photo-ionization and radiative recombination. Collisional processes become important at
higher densities. A source for many of our rates was Abel (1996), who collected up-to-date data for many of these processes.

For the ionization cross-sections we use the expression of Osterbrock (1989), which are linear sums of power laws and thus easily integrable.

\[
\begin{align*}
    r_{\gamma,i} &= \int_{\nu_T}^{\infty} \frac{4\pi J_\nu}{h \nu} a_\nu d\nu \text{ cm}^{-3}\text{s}^{-1}, \\
    a_\nu &= a_T \left[ \beta \left( \frac{\nu}{\nu_T} \right)^{-s} + (1 - \beta)(\nu/\nu_T)^{-2-1} \right], \nu > \nu_T.
\end{align*}
\]

where \( h \nu_T \) is the threshold energy, 13.6 eV for hydrogen and 24.6 eV and 54.4 eV for helium for the first and second ionizations, respectively. \( J_\nu \) is the mean intensity of the radiation field at frequency \( \nu \). \( a_\nu \) is the cross-section for photo-ionization as a function of \( \nu \). The constants are \( a_T = 6.3, 7.83 \) and \( 1.58 \times 10^{-18} \text{ cm}^{-2} \); \( \beta = 1.34, 1.66 \) and \( 1.34 \); \( s = 2.99, 2.05 \) and \( 2.99 \) for HI, HeI and HeII ionization respectively. With a fixed spectral shape we can evaluate the integrals and express the photo-ionization in terms of a rate that may be scaled by raising or lowering the whole radiation spectrum. We can also evaluate the heating rate.

\[
\text{heat}_{\gamma,i} = \int_{\nu_T}^{\infty} 4\pi J_\nu a_\nu d\nu \text{ cm}^{-3}\text{s}^{-1}.
\]

The recombination processes are detailed as follows: Collisional Ionization of hydrogen (Janev et al. 1987):

\[
\begin{align*}
    \text{H} + \text{e}^{-} &\rightarrow \text{H}^+ + 2 \text{ e}^{-}, \\
    r_{\text{ci},\text{HI}} &= \exp(-32.713967867 + \log_e(T \text{ (eV)}))(1.536556 \\
    &+ \log_e(T \text{ (eV)})(-5.73932875 + \log_e(T \text{ (eV)}))(1.56315498 \\
    &+ \log_e(T \text{ (eV)})(-0.2877056 + \log_e(T \text{ (eV)})(3.48255977 \times 10^{-2} \\
    &+ \log_e(T \text{ (eV)})(-2.63197617 \times 10^{-3} \\
    &+ \log_e(T \text{ (eV)})(1.11954395 \times 10^{-4}
\end{align*}
\]
Radiative recombination of hydrogen (Verner & Ferland (1996)). This rate differs only slightly from the rate given in Black (1981).

\[ \text{H}^+ + \text{e}^- \rightarrow \text{H} + \gamma. \]

\[
\begin{align*}
    r_{\text{rad},\text{HII}} &= \frac{7.982 \times 10^{-11}}{\sqrt{0.563615 T(K)}} \times (1 + 0.563615 \sqrt{T(K)})^{0.252} \\
    &\times (1 + 1.192167 \times 10^{-3} \sqrt{T(K)})^{1.748} \text{ cm}^{-6}\text{s}^{-1} \tag{2.46}
\end{align*}
\]

Collisional ionization of neutral helium (Janev et al. 1987):

\[ \text{He} + \text{e}^- \rightarrow \text{He}^+ + 2 \text{ e}^- \]

\[
\begin{align*}
    r_{\text{ci,Hel}} &= \exp(-44.09364886 + \log_e(T(\text{eV}))(23.91596563 \\
    &\quad + \log_e(T(\text{eV}))(10.7532302 + \log_e(T(\text{eV}))(3.05803875 \\
    &\quad + \log_e(T(\text{eV}))(6.79539123 \times 10^{-2} \\
    &\quad + \log_e(T(\text{eV}))(5.00905610 \times 10^{-3} \\
    &\quad + \log_e(T(\text{eV}))(2.06723616 \times 10^{-4} \\
    &\quad + \log_e(T(\text{eV}))(3.64916141 \times 10^{-6})))))))) \text{ cm}^{-6}\text{s}^{-1} \tag{2.47}
\end{align*}
\]

Collisional ionization of once ionized helium (Aladdin Database 1989 (Abel 1996)).

\[ \text{He}^+ + \text{e}^- \rightarrow \text{He}^{++} + 2 \text{ e}^- \]

\[
\begin{align*}
    r_{\text{ci,HeII}} &= \exp(-68.71040990 + \log_e(T(\text{eV}))(43.93347633 \\
    &\quad + \log_e(T(\text{eV}))(18.4806699 + \log_e(T(\text{eV}))(4.70162649 \\
    &\quad + \log_e(T(\text{eV}))(7.6924663 + \log_e(T(\text{eV}))(8.113042 \times 10^{-2} \\
    &\quad + \log_e(T(\text{eV}))(2.47) \text{ cm}^{-6}\text{s}^{-1} \tag{2.48}
\end{align*}
\]
Radiative recombination of once-ionized helium (Verner & Ferland (1996)):

\[
\text{He}^+ + e^- \rightarrow \text{He} + \gamma,
\]

\[
\begin{array}{l}
\frac{3.294 \times 10^{-11}}{0.253673 \sqrt{T} (1 + 0.253673 \sqrt{T})^{0.309} (1 + 1.649348 \times 10^{-4} \sqrt{T})^{1.691}}, \\
\frac{9.356 \times 10^{-10}}{4.841607 \sqrt{T} (1 + 4.841607 \sqrt{T})^{0.2108} (1 + 4.628935 \times 10^{-4} \sqrt{T})^{1.7892}}, \\
\end{array}
\]

\[ r_{rad, He} = \left\{ \begin{array}{ll}
T < 10^6 K, \\
T > 10^6 K, \\
\end{array} \right. \]  

\[ \text{cm}^{-6} \text{s}^{-1} \]  \hspace{1cm} (2.49)

Dielectric recombination of once ionized helium (Aldovandi & Pequignot 1973 (Black 1981)),

\[
\text{He}^+ + e^- \rightarrow \text{He} + \gamma,
\]

\[
r_{die, He} = 1.9 \times 10^{-3} T^{-1.5} \exp(-4.7 \times 10^5 / T) \\
(1 + 0.3 \exp(-9.4 \times 10^4 / T)) \text{ cm}^{-6} \text{s}^{-1}
\]  \hspace{1cm} (2.50)

Radiative recombination for doubly-ionized helium (Verner & Ferland (1996)):

\[
\text{He}^{++} + e^- \rightarrow \text{He}^+ + \gamma,
\]

\[
r_{rad, He II} = 1.891 \times 10^{-10} / (0.326686 \sqrt{T(K)})^{0.2476} \\
(1 + 6.004084 \times 10^{-4} \sqrt{T(K)})^{1.7524} \text{ cm}^{-6} \text{s}^{-1}
\]  \hspace{1cm} (2.51)

The set of rate equations,

\[
\frac{dn_{HI}}{dt} = -r_{\gamma, HI} n_{HI} - r_{\text{ci, HI}} n_{HI} n_e + r_{\text{rad, HI}} n_{HI} n_e,
\]  \hspace{1cm} (2.52)
can be posed in terms of ratios of the species number density to that of baryons.

\[ Y_i = n_i/n_b. \]

To find equilibrium solutions, we perform iterations on the equations with the derivatives set to zero.

\[ \frac{d n_{HeI}}{dt} = -r_{\gamma,HeI} n_{HeI} - r_{ci,HeI} n_{HeI} n_e + (r_{rad,HeI} + r_{die,HeI}) n_{HeI} n_e, \]

\[ \frac{d n_{HeII}}{dt} = -r_{\gamma,HeII} n_{HeII} - r_{ci,HeII} n_{HeII} n_e + r_{rad,HeII} n_{HeII} n_e, \]

\[ n_{HeI} = n_H - n_{HI}. \]

\[ n_{HeII} = n_{He} - n_{HeI} - n_{HeII}. \]

\[ n_e = n_{HI} + n_{HeII} + 2 n_{HeIII}. \]

\[ Y_e = (Y_H + 2 Y_{He}) - Y_{HI} - 2 Y_{He} - Y_{HeII}. \]

\[ f_{HI} = \frac{r_{ci,HI}}{r_{rad,HI} + r_{\gamma,HI}/(Y_e n_b)}. \]

\[ f_{HeI} = \frac{r_{ci,HeI}/(r_{rad,HeI} + r_{die,HeI}) + r_{\gamma,HeI}/(Y_e n_b)}. \]

\[ f_{HeII} = \frac{r_{ci,HeII}/(r_{rad,HeII} + r_{\gamma,HeII}/(Y_e n_b)}. \]

\[ Y_{HI} = Y_H/(1 + f_{HI}). \]

\[ Y_{HeI} = Y_{He}/(1 + f_{HeI}(1 + f_{HeII})). \]

\[ Y_{HeII} = Y_{He}/(1/f_{HeI} + (1 + f_{HeII})). \]

We do not follow the formation of molecular hydrogen, as it is not important except at high densities and low temperatures. It has also been shown (Shapiro and Kang (1987)) that the formation of H\(_2\) can only proceed efficiently under non-equilibrium conditions, because a fair supply of electrons is required to form the fragile intermediary, H\(^2\). Molecular hydrogen formation and cooling processes are only important for conditions similar to those in the dense Galactic ISM – which is not the focus of this work.

*Ultraviolet Flux Spectrum*
We use a spectrum that is given by a piecewise power-law, 
\[ J_\nu = J_{-21}((h\nu)/(h\nu_0))^{-\alpha}10^{-21}\text{erg cm}^{-2}\text{Hz}^{-1}\text{s}^{-1}\text{Sr}^{-1}. \]
We begin with the commonly assumed \( \alpha = 1.5 \) quasar spectrum which consistent with determinations such as that of Tytler et al. (1995) for the range 330 - 1300 Å at \( z \sim 3 \) (\( \alpha = 1.4 \pm 0.25 \)). Haardt & Madau (1996) performed a detailed examination of the propagation of quasar light through an absorbing and re-emitting medium. The result is a diminished flux above 13.6 eV (below 912 Å) and features in the spectrum associated with both photo-ionization and emission such as HeII Lyman-\( \alpha \). This motivates our spectral shape given in table 1.

The details of the spectral shape are subsumed into the photo-ionization rates for the important species, determined by integrals of the spectrum. Our rates are shown in figure 2.1. The spectral shape will have an important impact on the heating rates associated with the photo-ionization processes. The exact choice of spectral shape affects the ratio of the ionization rate to the commonly quoted flux at the Lyman edge, \( J_{-21} \). For our spectral shape, the HI photo-ionization rate of \( 7 \times 10^{-13}\text{s}^{-1} \) (Rauch et al. (1997b)) is produced by a flux of \( J_{-21} = 0.246 \).

The normalization we use is motivated by several determinations of the flux level. Haardt & Madau (1996) theoretically determined \( J_{-21} \sim 0.5 \) at \( z = 2 - 4 \) from the quasar luminosity function. Determinations based on the proximity effect give: e.g. Bechtold (1994), \( J_{-21} = 1.0 - 3.0 \) at \( z \sim 3 \), and Giallongo et al. (1997) \( J_{-21} = 0.5 \pm 0.1 \) at \( z = 1 - 4 \). Matching our exploratory simulations to the observed opacity of the medium gave \( J_{-21} = 0.3 - 0.5 \) at \( z = 2 - 3 \) for \( \Omega_\text{b}h^2 = 0.0125 \) (varying slightly with the cosmology). Bechtold notes that moving to newer slope estimates

Table 1: Ultraviolet Flux Spectrum.

<table>
<thead>
<tr>
<th>( h\nu ) range (eV)</th>
<th>( J_{-21} )</th>
<th>Slope, ( \alpha )</th>
</tr>
</thead>
<tbody>
<tr>
<td>13.6-24.6</td>
<td>0.5</td>
<td>1.5</td>
</tr>
<tr>
<td>24.6-54.4</td>
<td>0.2055302</td>
<td>1.0</td>
</tr>
<tr>
<td>54.4-544</td>
<td>0.025</td>
<td>1.1</td>
</tr>
<tr>
<td>544 &amp; up</td>
<td>0.0019764</td>
<td>1.5</td>
</tr>
</tbody>
</table>
$(\beta = 1.3 \text{ from } \beta = 1.7)$ for the column density distribution for clouds and other systematic effects lower her estimate by a factor $\sim 3$. $J_{-21} = 0.5$ is thus a well motivated choice at $z = 2 - 3$.

As no evolution is consistent with the data, we use a simple, fixed flux out to redshift $z = 15$, on the assumption that this is the redshift at which ionizing radiation sources first begin to make a significant contribution. We have investigated fluxes varying similarly to that of Haardt & Madau (1996) and different flux levels.

Figure 2.2 demonstrates the high degree of ionization expected at $z = 3$ for reasonable values of the UV flux. The curves demonstrate the inverse scaling of the neutral fraction with the flux level for low densities and where this starts to fail with the onset of significant collisional ionization. The assumption that photo-ionization alone is the important process is incorrect because shock heating brings the gas into the regime where collisional ionization is important ($T \gtrsim 100,000$K). Furthermore, this becomes the case sooner for hydrogen than helium. The assumption of a fairly uniform helium II to hydrogen I ratio $n_{HeII}/n_{HI}$ is thus slightly incorrect. The expected variation is largest for dense shocked gas ($n_0 \sim 10^{-3}$). We demonstrate instances of this in Chapter Four.
Fig. 2.1.— Ionization rates. The curves give photo-ionization rates for HI (solid), HeI (dashed) and HeII (dotted). The constant lines were used in our standard simulations. The thick curves are from Haardt & Madau (1996) and the thin curves were used in a few simulations.
Fig. 2.2.— Number of ionized species per baryon, for 4 different baryon number densities, $n_b$, at reshift $z = 3$. Each panel contains 12 curves. Solid lines are for HI, dashed curves for HeI and HeII and dotted curves for electrons. The top curve in each species set is for no UV flux, the middle with $J_{-21} = 0.2$ and the lowest with $J_{-21} = 0.5$. At $z=3$, average $n_b$ is $10^{-5}\,\text{cm}^{-3}$ (top left panel).
3.2. Cooling Processes and Temperatures

Our expression for the total energy includes a contribution from each ionized species equal to the energy required to ionize it from a neutral state. Thus we do not need to explicitly include a cooling process associated with collisional ionization to maintain a correct temperature. The cooling processes we must include are Compton cooling, bremsstrahlung, radiative recombination, dielectric recombination and line cooling. An overview of the processes is shown in figures 2.4 and 2.5.

The expression for Compton cooling, due to up-scattering of microwave background photons by electrons, is as follows.

\[ \Gamma_{\text{comp}} = n_e (5.65 \times 10^{-10} T_{\text{cmb} \cdot z=0} (1+z))^{4}(T - T_{\text{cmb} \cdot z=0} (1+z)) \text{erg cm}^{-3}\text{s}^{-1}. \] (2.65)

The present day temperature of the microwave background is \( T_{\text{cmb} \cdot z=0} = 2.735 \text{K} \).

Bremsstrahlung is the process whereby electrons radiate when moving with the electric field of ions. The cooling rate is (Kang & Shapiro (1992)).

\[ G_1 = \begin{cases} 0.79464 + 0.1243 \log_{10}(T)(n_{\text{HII}} + n_{\text{He II}}), & T < 3.2 \times 10^{5} \text{K} \\ 2.13164 - 0.1240 \log_{10}(T)(n_{\text{HII}} + n_{\text{He II}}), & T > 3.2 \times 10^{5} \text{K} \end{cases} \]

\[ G_2 = \begin{cases} 2.8792 + 0.4972 \log_{10}(T) n_{\text{He II}}, & T < 12.8 \times 10^{5} \text{K} \\ 8.8252 - 0.4960 \log_{10}(T) n_{\text{He II}}, & T > 12.8 \times 10^{5} \text{K} \end{cases} \]

\[ \Gamma_{\text{bremss}} = n_e 1.426 \times 10^{-27} \sqrt{T}(G_1 + G_2) \text{erg cm}^{-3}\text{s}^{-1}. \] (2.66)

Radiative recombination processes release a photon, thus cooling the gas with rates given with reference to the recombination rates above (2.46, 2.49 & 2.51).

\[ c_{\text{HI}} = n_{\text{HI}} n_e \Gamma_{\text{rad.HII}} (13.6E + k Te^{-0.0215964T^{0.270251}}) \]

\[ c_{\text{He I}} = n_{\text{He I}} n_e \Gamma_{\text{rad.HII}} (24.59E + k Te^{-0.0215964(13.6/24.59) T^{0.270251}}) \]

\[ c_{\text{He II}} = n_{\text{He II}} n_e \Gamma_{\text{rad.HII}} (54.42E + k Te^{-0.0215964(13.6/54.42) T^{0.270251}}) \]

\[ \Gamma_{\text{ci}} = n_e (c_{\text{HI}} + c_{\text{He I}} + c_{\text{He II}}) \text{erg cm}^{-3}\text{s}^{-1}. \] (2.67)
Fig. 2.3.— Recombination Cooling curve. Abel (1996) used the recombination rates from Ferland et al. (1992) but a used poor approximation to the cooling rate. We used the accurate fit shown.
where \( k = 1.38066 \times 10^{-16}\text{erg}\text{K}^{-1} \) is the Boltzmann constant and 
\( E = 1.60189 \times 10^{-12}\text{erg}\text{eV}^{-1} \). These rates are fits to the data given in
Ferland et al. (1992) (better than 4 \% error for \( 10^3 K < T < 10^7 K \)). It is suggested
by Abel (1996) to use the simple expression \( 1.04 kT \) to model the substantial extra
cooling due to the radiated kinetic energy of the recombining electron. This formula
is moderately accurate only for very low temperatures. Figure 2.3 shows the cooling
rates for HI. The drop off in the cooling compared to \( 1.04 kT \) reflects the much
lowered cross-section for recombination of fast moving electrons.

Recombination cooling makes a substantial contribution to cooling rates for
\( T \gtrsim 10^5 K \). The expressions given in Black (1981) are accurate over the range
\( 10^4 K < T < 10^5 K \) and tend to give the correct equilibrium results for \( T_{eqm} > 10^4 K \).
Below this temperature, the equilibrium temperatures are substantially over
estimated. This this would have important repercussions for simulations aimed
at the Helium Lyman-\( \alpha \) forest, which is dominated by quite tenuous gas with
equilibrium temperatures below \( 10^4 K \). Figures 2.4 and 2.5 show the impact of the
recombination cooling rate function graphically. Though some early simulations
employed the Black rates, most current workers use the accurate Ferland et al.
results.

Dielectric recombination cooling makes only a moderate contribution to the
cooling due to Helium at higher temperatures. This rate is fairly uncertain so that
the following represents an estimate only (Abel (1996)). The cooling rate is given by
(from 2.50),

\[
\Gamma_{\text{die}} = n_e n_{\text{He}I} \tau_{\text{die,He}I}(3.136 E + 24.59 E)\text{erg cm}^{-3}\text{s}^{-1}.
\] (2.68)

We use the hydrogen and helium line cooling rates given in Black (1981),
including the correction advocated by Cen (1992) of \( 1/(1 + \sqrt{T/10^5 K}) \). These
processes dominate the cooling at lower temperatures (\( T \sim 10^4 K \)) and high densities.
Fig. 2.4.— Cooling curves for different processes, for 4 different baryon number densities, $n_b$ (cm$^{-3}$), at reshift $z = 3$. The values are divided by $n_b^2$. The curves for the processes are labelled. The black solid curve is total |Cooling-Heating|, the cusp indicating the equilibrium temperature. The dashed curve uses the Black (1981) recombination formulae and the dotted curve uses the Abel (1990) approximation. $J_{-21} = 0.5$. 
Fig. 2.5.— Cooling curves for different processes, for 4 different baryon number densities, $n_b$ (cm$^{-3}$), at reshift, $z=3$. The values are divided by $n_b^2$. The curves for the processes are labelled. The black solid curve is total |Cooling-Heating|, the cusp indicating the equilibrium temperature. The dashed curve uses the Black (1981) recombination formulae and the dotted curve uses the Abel (1990) approximation. $J_{-21} = 0.5$. 
Figure 2.6 demonstrates the important cooling processes for gas of primordial composition. The curves shown are net cooling = cooling - heating. The top curves in each panel were computed with no UV flux, the next curve with $J_{-21} = 0.2$ and the lowest with $J_{-21} = 0.5$ erg cm$^{-3}$s$^{-1}$Hz$^{-1}$Sr$^{-1}$. The two UV fluxes shown bracket the range necessary to fit the observed Lyman-$\alpha$ forest opacities with a baryon density, $\Omega_b \sim 0.05$. The number densities shown correspond to the ranges typically found in objects at $z = 3$ that would produce detectable Lyman-$\alpha$ absorption. The curves are divided by $n_b^2$. This makes the curves effectively independent of $n_b$ for 2 body processes involving ions. Compton cooling and photo-ionization involve one baryon and a photon and line cooling involves neutral gas and thus neither scale in this way. Compton cooling is important for low densities and increasing redshift, as it scales as $(1 + z)^4$. The effect of the UV flux also drastically reduces the abundances of partially ionized species, acting to suppress line cooling. Line cooling is responsible for the humps in the curves between 10,000K and 100,000K. The dominant cooling process above 1,000,000K is Bremsstrahlung for high density gas and Compton cooling for lower densities. Due to our convention for total energy, we do not include collisional ionization as a cooling process, because it modifies temperature and not total energy. The curves including collisional ionization as a source of cooling are the same shape but are of moderately higher amplitude around 10,000K.

The heating due to the photoionizing radiation is proportional to $n_{\text{neutral}}$ for each species which is thus roughly proportional to $n_b^2$ if photo-ionization is the

\[
\begin{aligned}
l_{cH\text{I}} &= n_{H\text{I}} \times 10^{-19} e^{-1.18348 \times 10^5 / T}, \\
l_{cHe\text{I}} &= n_{He\text{I}} \times 10^{-17} e^{-0.1687 \times 10^5 / T}, \\
l_{cHe\text{II}} &= n_{He\text{II}} \times 10^{-17} e^{-0.3978 \times 10^5 / T}, \\
\Gamma_{\text{line}} &= \frac{n_e (l_{cH\text{I}} + l_{cHe\text{I}} + l_{cHe\text{II}})}{1 + \sqrt{T / 10^5}} \text{erg cm}^{-3} \text{s}^{-1}.
\end{aligned}
\]
Fig. 2.6.— Net Cooling curves for 4 different baryon number densities, $n_b$, at reshift, $z=3$. The values are divided by $n_b^2$ to demonstrate the scaling with this quantity for 2-body processes. The top curve is with no UV, the middle curve with $J_{-21} = 0.2$ and the lowest with $J_{-21} = 0.5$. At $z=3$, average $n_b$ is $10^{-5}$ cm$^{-3}$ (top left panel). Compton cooling is dominant in this panel.
dominant process. Ion related cooling processes also have an $n_b^2$ dependence and thus the interplay is a little subtle. For low densities ($n_b < 10^{-5}$ cm$^{-3}$), Compton cooling, proportional to $n_b$ alone, is large compared to the heating and drives low density gas to very low temperatures – as low as 1000K – even before the effect of expansion cooling. Expansion cooling suppresses low density equilibrium temperatures further. Cooling times become quite long at low densities, so that portion of the medium tends to have a memory of its temperature history. This is not important for HI but should be quite important for HeII absorbers – which are associated with low density material. Recombination times at the lowest densities are quite long, exceeding the Hubble time at $z \sim 5$; thus accurate abundance estimates require a non-equilibrium solution.

At high densities ($n_b \gtrsim 0.01$ cm$^{-3}$), efficient line cooling drags the temperature down to around $10^4 K$. The temperature will not drop below this value without additional cooling from metals or molecular hydrogen. neither of which are included here. Once the column of neutral hydrogen approaches $10^{18}$ cm$^{-2}$, the medium is optically thick to the ionizing photons. It is beyond the capability of current techniques to model radiative transfer self-consistently. For this reason, numerical results for high column objects can not be conclusive.

The peak temperatures in the medium occur in regions denser than the mean ($n_b \sim 10^{-4}$ cm$^{-3}$) at $z = 3$. These densities are also typical of the outskirts of dwarf galaxy type objects that are collapsing and merging so that the gas is often shock heated to quite high temperatures. The result is a large spread of temperatures, from a few times $10^4$ K to as high as $10^6$ K. The equilibrium temperatures discussed here provide a lower cutoff for the temperatures generated in the simulation and for temperatures expected to be observed in the IGM via Lyman-$\alpha$ observations.

In figure 2.7 we demonstrate how high-ionization acts to give an equilibrium temperature that is nearly independent of the amplitude of the ionizing flux at $z = 3$. At low densities the gas is highly ionized and the photo-ionization heating rate is a function of the neutral fraction multiplied by the flux. The neutral fraction
goes as density squared divided by the flux. The net result is a heating rate that is independent of the flux. The ionic cooling process that dominate are also substantially independent of the flux and thus the equilibrium temperature is also.

As the density rises, collisional ionization becomes important and then substantial recombination occurs. For neutral or collisionally ionized gas the photo-ionization heating rate is directly proportional to the ionizing flux level. At higher densities, however, shock heating is an important additional heat source. The curves have been plotted as line widths to facilitate comparisons with simulated absorption features in Chapter Four. The curves labelled 0.125 and 0.5 bracket the expected range of ultraviolet fluxes.

4. Gravity

The structure we see in the universe is a product of gravitational clustering. Local versions of the small fluctuations seen on the cosmic microwave background sky have evolved to generate the observed large scale structure. We are solving for small scale structure but the underlying process is the same. we need a solution for the Poisson equation.

$$\nabla^2 \phi = -4\pi G (\rho + 3p).$$

\(\phi\) is the gravitational potential, \(\rho\) is the density and \(p\) is the pressure. For relativistic particles \(p=\rho\). For the regimes of interest to us we can neglect physical pressure. Poisson’s Equation is a Newtonian local approximation to the general relativistic form, Einstein’s Field Equation that describes the universe (Peebles (1993)),

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

where \(R_{ij}\) and \(R\) are functions of \(g_{ij}\) the metric and \(T_{ij}\) is the stress-energy tensor. We can separate out the background cosmology and the local gravitational evolution
Fig. 2.7.— Line widths at temperature equilibrium as a function of density for different values of the ultraviolet flux at $z = 3$. The key feature to note is that there is almost no variation with changes in the flux until high densities.
and simply think of the simulations as evolving Newtonian volumes (with a simple modification for a non-zero cosmological constant).

There are two major approaches to solving gravity for particle codes: grid based methods (Monaghan and Lattanzio (1985), Evrard (1988), Couchman (1991)) and hierarchical tree codes (Hernquist and Katz (1989), Benz (1988)). Tree codes are extremely accurate but computationally demanding $O(N \log N)$ methods ($N$ is the number of particles). Grid methods are computationally fast, $O(N)$ methods (actually $O(N_{grid} \log_{2} N_{grid})$) but do not achieve the same level of resolution or accuracy because the grid size limits the resolution. Grid methods exhibit more flexibility in the implementation of different boundary conditions. A refinement to the basic PM (particle-mesh) grid based method is the calculation of local gravity with a direct sum: the P$^3$M algorithm (particle-particle-particle-mesh). A relatively new approach, the fast multipole method, is a gravity solver based on a refinement to the hierarchical tree codes which makes them methods of $O(N)$ (Greengard and Rokhlin (1986)). This last method has a high computational overhead and is thus suited to only very large numbers of particles and is not in general use. There is a limit to the force accuracy required for SPH-N-body codes given the intrinsic small errors involved in the SPH method. Accurate conservation of energy is of greater concern.

The initial code employed a 4 level multigrid solver. This was quite inefficient. We implemented a full-multigrid solver, based on the algorithm as presented in Press et al. (1992). We made several refinements to the algorithm that increased its speed and reduced the memory storage requirements. To this we later added a particle-particle correction which greatly increased the resolution possible (P$^3$MG). A thorough investigation of the errors of multigrid and the availability of a fast Fourier transform code (from Hugh Couchman) led us to switch to the latter technique. For highly clustered calculations the particle-particle section of the gravity would dominate the calculation time. This prompted us to develop a new technique that treats the particle-particle correction with a tree expansion that
dramatically improves computational costs.

Solving Gravity with Multigrid

The initial code was based upon the fourth order multigrid gravity solver of Monaghan and Varnas (1988) upgraded to full multigrid (Press et al. (1992)). The following sections describe the steps necessary to acquire the particle forces. The key flaw in the method is the difficulty in reducing the error arising from the grid.

1. Mass Interpolation

The mass of the particles is interpolated onto a grid using the $M_4$ basis spline. The $M_4$ kernel was chosen because the accuracy of the resulting force is highly dependent on the kernel choice. The higher order the basis spline, the better the accuracy. The $M_4$ kernel has a footprint of 4 particles cubed, making interpolation steps computationally expensive. Using the $M_4$ spline results in a grid derived rms error peaking at $\sim 5\%$ at a separation of 2 grid cells. A more typical rms error is $\sim 2\%$. This error can only be reduced by taking a higher order basis spline, resulting in much greater computational expense. The interpolation operation is common to all grid based solvers and is $O(N)$. We demonstrate the impact of modifying the interpolation scheme of force errors in figure 2.8. The errors for the $M_3$, $M_4$ and $M_5$ kernels are shown. The impact of these pairwise mesh force errors on the net gravitational force errors within a simulation are detailed in section 4.4.

2. Boundary Values for the Potential

The potential at the boundary of the calculation is evaluated using a multipole expansion (Binney and Tremaine (1987)). Only the moments up to the 6th are required for good accuracy. The grid must have a buffer of 4 cells empty of particles around the boundary for the multipole expansion to be sufficiently accurate at the boundary. This operation is $O(N)$ and only required for the multigrid solution in a non-periodic case.
Fig. 2.5.— Pairwise mesh force errors. This figure shows the effect of changing interpolation schemes (for multigrid). The top curve in each pair is error versus the mesh force and the lower curve is error versus a high resolution total force (approximated by $1/r^2$) which would include p-p corrections. The grey curves in the lower half give the average smoothed mesh forces. The grey curves in the upper panel give the difference between the mesh force and $1/r^2$. 

[Graph showing pairwise mesh force errors]
3. Full multigrid Solution for the Potential

The density and the boundary conditions are the initial data required to solve for the gravitational potential. We use a modified version of the full multigrid Poisson solver of Press et al. (1992). The algorithm was refined in the manner suggested by Monaghan and Varnas (1988) to make it fourth order accurate. This involves using a stencil that uses information from 18 points plus the Laplacian of the density field to eliminate the second order error term in the finite difference approximation to the Poisson equation.

\[
\sum w_{i,j,k} \phi_{i,j,k} = -4\pi G \left[ \rho + \frac{h^2}{12} \nabla^2 \rho \right]_{0,0,0}.
\]

(2.72)

where \(i, j\) and \(k\) may take the values 0, +1 or -1. The weights for the stencil are \(w_{0,0,0} = 4/(6h^2), w_{i,j,k} = 2/(6h^2)\) for two of \(i,j,k\) being zero and \(w_{i,j,k} = 1/(6h^2)\) for one of \(i,j,k\) being zero. \(\phi_{i,j,k}\) is the discrete representation of the potential at the point labelled \(i, j, k\).

The multigrid scheme is an iterative solver. It is efficient because we can inject the error in the solution information onto grids of every second grid cell of the previous grid recursively and thus solve for the low frequency components which are the slowest to converge. We use a top grid that is \(129^3\) grid cells. On the smallest \(3 \times 3 \times 3\) grid the solution is trivial as only one value is unknown. The solution is brought from low resolution grids back to the top grid with interpolation, which is appropriate because we are dealing with a small correction which does not contain the highest frequencies. This called a V-cycle; travelling from the top to the smallest grid and bringing the correction. V-cycles are repeated until the solution has converged.

The algorithm tests for convergence by assessing the rms error divided by the source term and assuming convergence when this is below 0.001. Typically this would take 6-8 iterations. During a calculation we can use the previous step as a first guess for the solution, in which case only 2-4 iterations are required. Iterating
to higher accuracy is pointless because of the intrinsic error associated with the mesh interpolation.

The multigrid solution procedure is $O(N_{grid} \log_2(N_{grid}))$. Typical grid choices are larger than the number of particles, by a factor of 2 in each dimension. This tends to result in balanced computational effort. For our calculations of 80 particles spacings across, we employed a $(128 + 1)^3$ grid. The method is restricted to powers of 2 times the innermost grid width-1 (here 3) plus one. \(i.e.\)

$n_{top} = (2^{\text{levels-1}}(n_{\text{bottom}} - 1)) + 1$ in each dimension.

4. Force Determination

The gravitational force at the grid vertices is calculated using the fourth order finite-difference scheme of Monaghan and Varnas (1988). This scheme uses the same principle as the scheme above to remove the second order error term using 18 points and the derivative of the density.

5. Force Interpolation

The gravitational force is interpolated back onto the particles using the \(M_4\) kernel. By using the same kernel to interpolate to and from the grid we ensure exact momentum conservation and no self forces (to within round-off error, Hockney and Eastwood (1988)). Verification of this fact provided a basic test of the code.

4.1. Particle-Particle Correction

Particle-particle (P-P) corrections allow a particle-mesh code to achieve high force resolution and also allow greater accuracy. The particle-particle correction step involves a local addition to the gravitational mesh force on the particles via a pair-wise correction that brings the gravitational force assigned from the mesh (PM) to the desired force, which can be of arbitrarily higher resolution than the mesh force. The correction is required out to a particle pair separation for which the mesh force becomes equal to a $1/r^2$ force-law. This is between 3 and 4 grid cells for the
multigrid with the $M_4$ kernel. The correction is an $O(n^2)$ procedure, where $n$ is the number of particles in each local region requiring the correction. The practical limit on the use of particle-particle corrections occurs when the computation for the mesh and the correction is similar. If a coarse mesh is chosen relative to the number of particles the computation for the correction will become large. This will also be the case for highly clustered $N$-body simulations. Therefore, the cost ratio is chosen so that the computational costs are comparable towards the end of a calculation when clustering is greatest.

For the particle-particle, particle-multigrid code ($P^3MG$), we chose the desired force to be the analytically calculated gravitational force between two spherical density peaks with $M_4$ kernel distributions. This was appealing because it perfectly matched the gas (SPH) density distribution. The mesh grid force was estimated with a large monte carlo sample of pairwise forces calculated by the mesh code. The P-P correction consists of adding the difference out to a radius of 3 grid spacings. At this separation, the difference was reduced to less than 0.5% and was neglected in favour of reduced computation. The P-P computational work goes as $r_{corr}^6$, where $r_{corr} = 3$ is the radius of the correction in grid spacings. We demonstrate the performance of the $P^3MG$ gravity code in comparison to the newer Tree$P^3M$ code in the following sections.

### 4.2. Fast Fourier Transform for Gravity

The fast Fourier transform (FFT, see *e.g.* Press *et al.* (1992)) provides an efficient way to solve the Poisson Equation for gravity. The basic principle of the fast Fourier transform is that the non-local information transfer required to produce the transform from the original data can be performed in $\log_2(N)$ steps each of $O(N)$. The final result is an $O(N \log_2(N))$ method for inverting a problem that would otherwise be $O(N^2)$ to solve by brute force. We use updated versions of the FFT code of Couchman (1991) and this discussion closely follows Couchman (1995).
Taking the Fourier transform of Poisson's equation renders the solution simple,

\[-k^2 \hat{\phi}(\mathbf{k}) = -4\pi G_{\text{grav}} \hat{\rho}(\mathbf{k}).\]  

(2.73)

where the \( k \) refer to independent Fourier modes. We have used a carat to indicate the Fourier transform. More generally the solution is expressed,

\[\hat{\phi}(\mathbf{x}) = \int d\mathbf{x}' \rho(\mathbf{x}') G(\mathbf{x}' , \mathbf{x}).\]  

(2.74)

with \( G(\mathbf{x}' , \mathbf{x}) = G_{\text{grav}} / |\mathbf{x} - \mathbf{x}'| \) for an unsmoothed force law. \( G \) is known as the Green's function. This is a convolution, so the solution is expressed in Fourier space as,

\[\hat{\phi}(\mathbf{k}) = \hat{G}(\mathbf{k}) \hat{\rho}(\mathbf{k}).\]  

(2.75)

The Green's function may be shaped with small modifications to reduce errors. This is most easily achieved by convolving a smoothing function in with the Green's function. Thus a highly compact (and fast) interpolating kernel may be used to put the particle information on the initial grid and smoothing may be performed to reduce the error associated with the low order without the high computational cost of high order interpolation (cf. section 4.). The kernel employed in the Fourier mesh code is the \( M_3 \) spline in the terminology of Schoenberg (1973), more commonly referred to as TSC (triangular shaped cloud) in the potential solver literature,

\[W_{\text{TSC,1D}}(x) = \begin{cases} 
\frac{3}{4} - x^2 & |x| \leq \frac{1}{2}, \\
\frac{1}{2} \left( \frac{3}{2} - |x| \right)^2 & \frac{1}{2} \leq |x| \leq \frac{3}{2}, \\
0 & |x| > \frac{3}{2}.
\end{cases}\]  

(2.76)

In three dimensions, a product of 3 one-dimensional kernels, one in each dimension, is used.

The improvement possible to pairwise mesh force errors through smoothing is
demonstrated visually in figure 2.9. Errors from ~ 2 cell separations are redistributed to larger and smaller distances. This is extremely advantageous because errors at ~ 2 cells separation are the most tenacious. Errors at smaller separations are rendered unimportant in comparison to the accuracy of the particle-particle scheme. Assuming one is used, because the particle-particle correction contributes the bulk of the force. At larger separations, the number of particles rises and the net force error is diminished by $\sqrt{1/N}$.

A second source of improved accuracy is the ability to optimize the Green's function to minimize both variations as the particles are moved relative to the grid points and variations in the forces with direction relative to the principle grid directions. These are the chief forms which the particle-mesh errors take and may not be addressed with multigrid schemes except through higher order interpolation. Translational and directional invariance of the force is important if a particle-particle correction is to be accurate. This is achieved by minimizing the square of the deviation of the mesh force, $F(x, x')$ from the desired force $R(x)$ with respect to the Green's function (Hockney and Eastwood (1988)).

$$Q = \frac{1}{V} \int_V \int_V dx \, dx' (F(x, x') - R(x))^2.$$  \hfill (2.77)

The desired force employed is the force between two density clouds with shape (labelled $S_2$ by Hockney and Eastwood (1988)).

$$\rho(r) = \begin{cases} \frac{48}{\pi a^4} (\frac{a}{2} - r) & r < \frac{a}{2}, \\ 0 & r \geq \frac{a}{2}. \end{cases}$$  \hfill (2.78)

The softening, $a$, is referred to hereafter as the Soft parameter. The transform of the desired force, $\tilde{R} = -i k \tilde{S}_2 / k^2$, given in terms of the Fourier transform of $S_2$, $\tilde{S}_2$. The Soft parameter must exceed 2 and generally 3.2 or better is required for reasonable accuracy. It is advantageous not to have too large a value for this parameter, or the force will converge slowly to $1/r^2$, making particle-particle correction computationally expensive. The value of Soft that matches the $S_2$ force
Fig. 2.9.— Pairwise mesh force errors. This figure shows the benefit of smoothing, given in brackets for each case (MG multigrid, $M_4$ kernel; PM Fourier mesh, TSC kernel). The top curve in each pair is error versus the mesh force and the lower curve is error versus a high resolution total force (approximated by $1/r^2$) which would include p-p corrections. The grey curves in the lower panel give the average smoothed mesh forces. The corresponding grey curves in the upper panel give the difference between the mesh force and $1/r^2$. 
to that between a pair $M_4$ kernels with smoothing lengths of unity is $a = 2.94$.

The mesh force can be expressed in terms of convolutions with interpolating, differencing and sharpening operators. In Fourier space the transforms of the interpolation function, $\hat{U}(k_n)$, the Green’s function, $\hat{G}(k)$ and the difference operator $\hat{D}(k)$ combine to express the solution for the mesh force:

$$ F(x, x) = \frac{1}{V} \sum_k \hat{D}(k) \hat{G}(k) \sum_n \hat{U}(k_n) e^{i k_n \cdot x} \sum_m \hat{U}(k_m) e^{-i k_m \cdot x}. \quad (2.79) $$

Minimizing the deviation gives the solution,

$$ \hat{G}(k) = \frac{\hat{D}(k) \cdot \sum_n \hat{U}(k_n) \hat{R}(k_n)}{|\hat{D}(k)|^2 \left[ \sum_n \hat{U}^2(k_n) \right]^2}. \quad (2.80) $$

The effect of Green’s function shaping in Fourier space is to modify the form of the differential equation in real space. It might technically be possible to employ modified differential or finite difference operators with the Multigrid method but it would be very complex and most likely unworkable.

The fast Fourier transform method is most effective for periodic boundary conditions. To solve for free boundaries, one must employ a technique whereby a single corner, $1/8^{th}$ of the volume contains active mass and the remainder is left empty. The Green’s function is is periodic but gives the correct force in the active corner of the mesh. The point mass Green’s function for this case is,

$$ G(x, x) = \frac{1}{\sqrt{(x' - x)^2 + (y' - y)^2 + (z' - z)^2}} \begin{cases} \frac{L}{2} \leq x' - x, y' - y, z' - z \leq \frac{L}{2} \end{cases}. \quad (2.81) $$

The cusp at $L/2$ never affects the solution.

The cost for an isolated solution is extra computation, typically a factor of $\sim 4$ over a periodic case with the same active grid volume. It is not a factor of 8 because
the empty cells mean that some sweeps of the transform need not be done.

We need only generate the Green's function transform once, at the start of the calculation when we select the smoothing, periodic grid sizes and so forth. Symmetries in $G$ can be used to reduce the storage required enormously from the full $(2N_{grid})^3$ for the non-periodic case.

We take advantage of the highly developed fast Fourier transform particle-mesh routines of Couchman and collaborators (part of the freely available AP$^3$M code, Couchman (1991)). Couchman has recently made available more efficient versions of his FFT routines. The code is well tested and highly accurate. The AP$^3$M code is distributed as a periodic solver, so the primary grid is periodic. Isolated grids are used for refinements to improve computational efficiency. The only modification the code required for our use was making an isolated grid the primary grid.

After incorporating the routines within our SPH code, we made comparisons with multigrid, specifically tests of force accuracy on the grid. In next section we discuss the inclusion of particle-particle forces, our refinements to that basic scheme and force accuracy tests. In section 4.7, we describe tests of the complete code.

4.3. Trees for Particle-Particle

Particle-particle correction is $\mathcal{O}(N_{corr}^2)$. $N_{corr}$ refers to the typical number of particles needing gravity force corrections. For most mesh schemes, the correction required becomes half the mesh errors at a separation of $r_{corr} = 2.4 - 3$ gravity cells. For uniform particle distributions, $N_{corr} \sim 58 - 113$. Using a simple link list over $(3r_{corr})^3$ cells, 373 – 729 particles would have to be checked against each other to see if they are within this distance. $<N_{corr}^2>$ becomes large with heavy clustering. Couchman (1991) solved this problem with adaptivity: The AP$^3$M code lays down additional meshes in regions experiencing high clustering. This refinement process continues within the new meshes until the computational cost of the final particle-particle step is brought under control. The code is large and
complex. A simpler method, which is more amenable to parallelization, is to use a
tree method for the correction. Tree methods are intrinsically $O(N_{corr} \log_2 N_{corr})$
with more computational overhead than a simple sum. Meshes are the fastest way
to accumulate the large scale force. Using a tree for the correction overcomes the
clustering problem. The tree structure we employ is very similar to that developed
for fast neighbour finding for the SPH hydrodynamical part of our calculations
(section 4.5.).

Concentric shells of constant thickness make equal contributions to the
gravitational force at the central point. At larger radii the details of the mass
distribution are less important, providing the basis of the tree method. The accuracy
of the force contributions is constant for a fixed value of the opening angle $\Theta = d/r$,
where $d$ is the size of the mass clump whose details are neglected and $r$ is the
distance to the clump. The correction force begins to drop much faster than $1/r^2$
near $r_{corr}$.

We set up the correction by generating spatial monopole trees in root cells with
width $d_{cell} = 1.5$ gravity cells. Spatial trees divide the particles into eight subgroups
each time with simple division of the volume in half in each dimension. We go out
two root cells to achieve a correction distance of 3 gravity cells. Using a monopole
tree means the leading errors are in the quadrupole. A quadrupole tree correction
is straightforward, requiring derivatives of the correction force and storage of the
quadrupole moments at all tree levels. The extra storage and computation involved
are not warranted and the monopole method is sufficiently fast and accurate for a
correction.

Each root cell can point to a particle or a data structure containing a mass
and centre of mass position for itself and pointers to 8 possible subcells (12 x 4
bytes). Each subcell pointed to may be a particle or a new data structure with 8
further subcells. When a subcell contains two particles, it is not subdivided but a
link points to the second particle instead. This modification was performed because
dark matter particles and gas particles can sit on top of each other. A standard tree
would subdivide indefinitely attempting to separate the two.

Of the $5^3$ root cells required for the correction, the outer layer of 98 are used directly and not tested against the opening angle criterion. A particle at the edge of the inner root cell would be 2.25 gravity cells away from the centre of the nearest outer root cell, implying an opening angle of $\Theta_{\text{cut}} = 2/3$, our standard value. A pathological particle arrangement could give $\Theta = 1$, which is still reasonable. The typical case is very much better. Additionally, the correction is smallest for these cells as noted above. 54 outer cells are excluded by being more than 3 gravity cells from the closest edge of the root cell. The remaining outer cells' information is used to create a basic list of masses and positions for the correction for each particle in the central cell, which can be summed over quickly and efficiently.

The remaining 27 root cells are used directly if they satisfy the opening criteria $\Theta = d/r < \Theta_{\text{cut}}$ for each central cell particle or opened and their subcells tested. The $d$ used is the cell size and $r$ is the distance from the particle to the cell centre of mass. Particles have zero size and always pass the test. If care is not taken, a pathological case of a particle in one corner and the centre of mass in the other could occur because $\Theta_{\text{cut}}$ is not less than $1/\sqrt{3}$. The error would entail the particle not opening its own cell. We always open the cell in which the particle resides and avoid this. The resulting code is efficient and short if coded recursively. The force summation step is identical to a direct particle-particle summation, where the “particles” summed are possibly aggregates rather than single particles.

The testing was done on the cluster comparison initial condition described in section 4.7. with 70,000 dark and 70,000 gas particles, including a nested high resolution region. The resulting method is very fast, as shown in table 2. The numbers are rough because the timing output was at slightly different redshifts to those shown. The runs were performed on a DEC alpha EV5 at 250 MHz. The direct particle-particle code (labelled pp) used a larger grid ($128^3$ mesh) to reduce the cost of the p-p. Build and Tree-pp refer to the time to build the tree and to do the force summation respectively.
4.4. Force Errors and Method Comparison

There are two potential sources of error in a particle-particle particle-mesh code. In weakly clustered simulations, such as of very large scale structure, it may not be necessary to use approximations to calculate the particle-particle force in a reasonable time. For small scale simulations, such as Lyman-α, the clustering is extreme and we move to approximate schemes. In practice, the mesh is the greater source of errors. In figure 2.8 we demonstrated the impact of higher order interpolation. In figure 2.9 we demonstrate the improvement in mesh force errors under smoothing and in figure 2.10 we compare ‘best’ versions of three mesh force methods.

The pairwise mesh error is not systematic but random about the mean. A given particle feels a mesh force that is the result of the contributions of many particles. thus the net error for moderate to large separations is small due to a Poisson type random error process reducing the error according to $\sim 1/\sqrt{N}$. The number of particles in a radial bin is proportional to $4\pi r^2 \Delta r$ so the net error is effectively reduced by a factor $\sim 1/r$ over the pairwise error. Thus even though the pairwise error is larger at 4 grid cells separation for the smoothed TSC kernels in figure 2.9, this effect makes it less important than the error at $\sim 2$ cells. The smoothing shown for the multigrid case had to be carried out by extending the interpolation region (at 4 times the computational cost), and is thus not practical but purely illustrative.

The mesh force generation schemes we employ are known as momentum conserving schemes, so that momentum is conserved correctly to within roundoff.

Table 2: CPU seconds per step for force correction.

<table>
<thead>
<tr>
<th>Redshift</th>
<th>Build</th>
<th>Tree-pp 64³</th>
<th>pp 128³</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.0</td>
<td>0.2</td>
<td>13</td>
<td>2</td>
</tr>
<tr>
<td>2.0</td>
<td>0.3</td>
<td>18</td>
<td>18</td>
</tr>
<tr>
<td>0.5</td>
<td>0.3</td>
<td>38</td>
<td>199</td>
</tr>
<tr>
<td>0.0</td>
<td>0.3</td>
<td>49</td>
<td>390</td>
</tr>
</tbody>
</table>
error. When a particle-particle correction force is added, the contribution to the total force by the mesh force below 1 grid cell separation tends to be very small. This means that large errors in this regime have little impact on the percentage pairwise total force errors that include the correction. In the figures this is indicated by the second lower curves which give the pairwise force error when compared to $1/r^2$. This would seem to be an advantage of the energy conserving scheme, labelled $E_{\text{con}}$ in figure 2.10. This scheme does not finite difference the potential, but uses the gradient of the interpolating kernel to directly convert the potential into a force, which gives better energy conservation but sacrifices momentum conservation. This leads to self-forces: isolated particles moving in small orbits due to their own mass. This is totally unphysical and thus the energy conserving scheme is to be avoided.

The maximum acceptable difference between the $1/r^2$ force and the mesh force, indicated on the figures, provides the criterion for the choice of the range of the particle-particle correction force. The computational cost of this correction goes as the range to the sixth power. For the schemes presented, the correction required becomes half the rms error around 3 grid cells for the PM. Soft=3.5 scheme and 3.5 grid cells separation for the multigrid. $M_4$ scheme indicating the p-p correction for the latter would require much more work.

We present a comparison of the complete gravity solvers in figure 2.11, derived from a (code force)/(direct-sum force) comparison on the late stages of a Lyman-α simulation. The forces errors shown are smaller than those discussed for the grids alone, not only because of the addition of particle-particle corrections, but also because the forces result from summations over large numbers of particles and thus the error tends to reduce in according to $\sqrt{N}$ compared to the pairwise force errors. The tail of the distribution is mostly associated with particles with small forces. The percentage error is large, but the absolute errors are small and the contribution to the numerical energy conservation error from such particles is insignificant.

Standard particle-particle correction is accurate to within roundoff error, but is too computationally expensive, so must be abandoned. The multigrid + p-p force
Fig. 2.10.— Pairwise mesh force errors. This figure compares "best" versions of different schemes (MG multigrid, $M_4$ kernel; PM Fourier Mesh, TSC kernel, softening 3.5: $E_{\text{con}}$ energy conserving scheme – see text). The top curve in each pair is error versus the mesh force and the lower curve is error versus a high resolution total force (approximated by $1/r^2$) which would include p-p corrections. The grey curves in the lower panel give the average smoothed mesh forces. The corresponding grey curves in the upper panel give the difference between the mesh force and $1/r^2$. 
errors are clearly uncomfortably large. This is entirely due to the large intrinsic mesh errors for the multigrid method. The force errors for particle-mesh derived from Fourier methods plus standard particle-particle correction are superb. To reduce the computational cost we have use the tree-particle-particle correction detailed in the previous section which is $O(N_{\text{sum}} \log_2 N_{\text{sum}})$ rather than $O(N_{\text{sum}}^2)$ for the correction summation. From the figure, the choices $\theta=0.67$, Cell=1.5 and soft=3.5 are clearly good (where $\theta$ and Cell refer to parameters of the tree-pp method and Soft is the Green's Function softening for the particle-mesh routine). They are slightly more computationally expensive (~20%) than the other options. Our production code thus has 94% of its particles with less than 1% force errors. The mean force error is 0.249% and the rms force error is 0.410%.

4.5. Local Trees for Neighbour Searching

Each particle in an SPH calculation generally interacts with of order 50 neighbours. Neighbour searching is the step which converts SPH from an $O(N^2)$ method to a $O(N \cdot N_{\text{neighbours}})$ method. This number of interactions is generally larger than Eulerian codes (minimally 6 interacting cells). and suffers from the necessity to locate the neighbours rather than just look at the next cell. It is important that this step be performed efficiently.

Linked lists (Monaghan (1985a)) are a very efficient means of finding neighbours for simulations with constant $h$ or where the variations in particle number density and $h$ are small. The basic procedure is to lay down a grid and bin the particles. Each cell has a pointer to the first particle in a linked list of all the particles within that cell. This method is very efficient when the cell sizes are comparable to the typical interaction distance.

For variable $h$ calculations, the simplest modification to the standard procedure is to construct a linked list with cell size equal to the range of interaction of the particle with largest $h$. For applications involving gravity, many particles can clump
Fig. 2.11.— Gravity force error distributions, differential (lower) and cumulative (upper curves). Direct standard p-p correction is exact to within roundoff error, so the errors given by the solid curve and the dashed curve are due to the mesh force. The mesh force is given by Fourier methods (all curves but dashed) or multigrid (dashed). $\theta$ and Cell refer to parameters of the tree-pp method (dotted, dot-dash and dotted-dashed curves). Soft refers to the softening radius for the Fourier mesh force, which is 3.2 if not otherwise stated.
within one linked list cell, while some cells are nearly empty, effectively reducing the calculation to $O(N^2)$. The typical interaction distance is a fraction of the cell size. A linked list uses a top array to point to the first particle in each cell and an additional integer per particle to point to the next particle in the list.

Because of the above problem and due to their intrinsic flexibility for irregular particle configurations, many authors use a tree code to locate neighbours in such situations (e.g. Hernquist and Katz (1989), Benz (1988), Evrard (1988)). If the gravity is also calculated with a tree code then the tree structure will already exist. Hernquist and Katz (1989) note, however, that the tree search is 3-5 times slower than a linked list for constant $h$.

There exists room for improving the linked list method. For example, Benz (1993) has suggested that instead of having fixed cell boundaries, the particles may be sorted in space in each coordinate direction and the cell boundaries set in each direction so that each cell contains a similar number of particles. Instead of looping over cells as with linked lists, the particles are treated in order of decreasing $h$ such that no particle ever interacts with a particle of greater $h$. This allows use of the symmetric nature of inter-particles forces to reduce computation without introducing the possibility of treating the same pair twice. The disadvantage of sorting and ranking methods is the slow speed of sorting and the fact that you may have to search more than one cell in a given direction to reach all interacting pairs.

We employ a technique similar to that developed by Theuns and Rathsack (1992) for a parallel SPH code. These authors laid down two grids with the second having half the spacing of the first. Particles are assigned to the largest grid they need for the interaction distance to still be one cell. This technique is limited because laying down more high resolution grids begins to consume large amounts of memory. We have developed the technique to use dynamic storage in a similar manner to a tree-code so that the subgrids exist only in larger cells which contain particles that can make use of them. We have made this technique recursive so that each subcell may in turn lay down new subcells. The total storage is $O(N)$ because the extra
grids follow the particles themselves.

The technique assigns particles to the upper-most grid (a regular array) if their interaction distance, \( d = 2h \), is greater than one half the grid spacing, determined by twice the largest \( d \) value for all gas particles. If \( d \) is less than half this value and no subgrid exists, a subgrid of 8 cells of half the top cell size is generated and a pointer to the subgrid (stored in a array of subgrids) is allocated to the grid cell. If a subgrid already exists, it is used. If \( d \) is still less than half the subgrid spacing, a new subgrid will be allocated and this process continues recursively. 8-9 levels of recursion are typical for a \( d \) range of \( \sim 260 \). Each subgrid requires ten 4 byte variables. The first points to the list of particles for the cell, 8 are pointers to possible further subgrids and one is a logical variable indicating if any subgrids are occupied. A standard tree needs more subgrids than there are particles, depending what the typical level of recursion is. For this algorithm there is less than half as many subgrids as particles because each subcell contains several particle. The maximum storage requirement for the linked list and root cells and sub cells can be estimated as seven 4 byte variables per particle. In our code we use work arrays also used by the gravity solver, so that no new storage is required.

We have tested this method on highly clustered simulations. The cut in CPU time is enormous for heavy clustering compared to a direct linked list. Most CPU time for a highly clustered case with a standard linked list is spent testing particle pair separations to see if they need to interact. Table 3 compares CPU costs for the cluster comparison calculation described in section 4.7. In this simulation a very rich, relaxed cluster is formed by \( z = 0.1 \).

Levels is the nested levels allowed (used), \( N_{\text{interactions}} \) refers to the interactions necessary and \( N_{\text{pairs}} \) is the number of pairs tested. The squareness of grids (rather than being spherical) combined with their finite volume (3 cells cubed) limits the best possible ratio of these numbers to be \( \sim 7 \). The simulation included \( N = 74127 \) gas particles ("None" corresponds to \( N^2 \)). It can be seen that as clustering advances, the standard linked list draws towards the \( N^2 \) case. Part of the reason for the
cost of standard linked-lists is the size of the largest interaction length. Artificial upper limits on this length will improve the speed a lot with the cost of rendering gas particles in voids non-gaseous with no SPH interactions. In the \( z = 0.1 \) case above, \( h \) ranged from \( 1/20^{th} \) to 8 times the initial spacing, comoving. The number of interactions required remains roughly constant at 17 pairs per particle, implying an average of 35 neighbours.

4.6. Particle Setup

Setting initial conditions for SPH simulations can be relatively straightforward. Given a distribution of density, temperature and other physical properties, particle properties such as mass and internal energy need to be assigned. The standard assignment procedure is to subdivide the total volume of the simulation into regions associated with each particle and to assign all mass, energy and so on contained within that region to the related particle. This assumes that the particle positions are already determined. This will result in varying particle mass (see Monaghan (1992)).

A second procedure is to use a fixed mass for each particle and to position the particles in such a way as to reproduce the desired density distribution. This approach is not as straightforward unless the initial condition desired is a small

<table>
<thead>
<tr>
<th>Levels</th>
<th>Redshift</th>
<th>( N_{\text{interactions}} )</th>
<th>( N_{\text{pairs}} )</th>
<th>CPU sec</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8.0</td>
<td>1242590</td>
<td>149528282</td>
<td>25.31</td>
</tr>
<tr>
<td>( \infty ) (3)</td>
<td>8.0</td>
<td>1242590</td>
<td>25602870</td>
<td>5.7</td>
</tr>
<tr>
<td>1</td>
<td>2.0</td>
<td>1256147</td>
<td>495796816</td>
<td>101</td>
</tr>
<tr>
<td>( \infty ) (6)</td>
<td>2.0</td>
<td>1256147</td>
<td>29688079</td>
<td>7.4</td>
</tr>
<tr>
<td>1</td>
<td>0.5</td>
<td>1245831</td>
<td>1437815035</td>
<td>272</td>
</tr>
<tr>
<td>( \infty ) (8)</td>
<td>0.5</td>
<td>1245831</td>
<td>37341204</td>
<td>9.6</td>
</tr>
<tr>
<td>1</td>
<td>0.1</td>
<td>1255990</td>
<td>1936987811</td>
<td>350</td>
</tr>
<tr>
<td>( \infty ) (8)</td>
<td>0.1</td>
<td>1255990</td>
<td>37672039</td>
<td>10.4</td>
</tr>
<tr>
<td>None</td>
<td>Any</td>
<td></td>
<td>5494812129</td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Neighbour Finding CPU Costs.
perturbation away from a uniform case. If the initial variation is large and the simulation in more than one dimension, the method is difficult. A one-dimensional algorithm is described in appendix D.

For cosmology, we can always commence the calculation at a time when the initial perturbations are small. We use a scheme called the Zel’dovich approximation (Zel’dovich (1970)), derived from linear theory to generate our initial conditions. Given a density perturbation distribution, $\delta(r) = \delta \rho(r)/\bar{\rho}$, we find the displacements and velocities for particles from the regular lattice that recreates this field self-consistently. The linearized equations for the growth of perturbations give us the peculiar velocities, $\mathbf{v}$ via.

$$\frac{d\delta}{dt} = \frac{1}{a} \nabla \cdot \mathbf{v}. \quad (2.82)$$

In the linear regime $\delta(r, t) = D(a) \delta_0(r)$, where $a(t) = 1/(1 + z)$ is used as a time variable and $D(a)$ is the amplitude of the fluctuations normalized to $D = 1$ at $a = 1$. Using the absence of vorticity in the initial velocity field, we can easily invert equation 2.82 to find $\mathbf{v}(r, t) = D(a) \mathcal{v}_0(r)$. The displacement $s$ is a time integral of $\mathbf{v}$, which in linear theory just means $\nabla \cdot s(r, t) = D(a) \delta_0(r)$. This is discussed in more detail in Chapter Three.

Initially small perturbations do not guarantee that the calculation will be correct, because higher order moments of the density distribution are very sensitive to the setup method (Scoccimaro (1997)). Scoccimaro concludes that higher order approximations or commencing the calculation many expansion factors prior to when results are measured is sufficient to overcome this problem. We are not concerned with highly accurate large scale measures of clustering but nevertheless we begin our calculations at redshift, $z = 60$, when typical fluctuations within our calculation are at the few percent level.

4.7. Standard Test Problems
Once an SPH code has been written, the code is applied to standard problems to test accuracy. The normal procedure is to run a few one dimensional cases to test the hydrodynamics and a spherical collapse problem in three dimensions to test the gravity. The desire for rigorous testing restricts the choice of test cases to those with analytical solutions and to those for which results are available from other stringently tested codes.

Multi-dimensional tests are essential as there are aspects of the hydrodynamics which can not be tested in one dimension. Monaghan (1988) demonstrated that while penetration in colliding streams is difficult in one dimension, in two dimensions particles may slip around each other. Similarly, instabilities in the direction perpendicular to the shock are only manifested in two or three dimensional simulations. We perform our basic hydrodynamical tests in three dimensions to ensure they are a robust test of the code. We have simulated the standard Sod’s shock tube problem (Monaghan and Gingold (1983)) in three dimensions. We also introduce a new test, an adiabatic Gaussian wave that breaks and forms a shock. The evolution of the wave is analytically predicted and provides an indicator of the diffusion present in the numerical scheme. The results are shown below.

Other shock type tests in the literature include colliding streams (Monaghan (1989)) and strong piston-driven shocks (Hernquist and Katz (1989)). More involved hydrodynamical test problems to which SPH has been applied include interacting spherical blast waves (Steinmetz and Müller (1993)), the evolution of pressureless rings orbiting a point mass (Monaghan (1989), for which SPH is known to perform poorly) and the onset of Rayleigh-Taylor instabilities (Monaghan (1992), for which SPH performs well). The extensive literature of SPH tests makes us confident in the robustness of the method.

Three-dimensional tests designed for gravity include the orbits of binary particles (a very stringent test, Monaghan and Lattanzio (1985)), self-similar adiabatic expansion and self-gravitating collapse (Evrard (1988)) and the equilibrium profile of gas polytropes (Nelson and Papaloizou (1993)). We have
tested the gravity-only aspects of the code with a study of the force errors and the self-gravitating collapse test, making comparisons among the techniques discussed.

Cosmological simulations are very complex and difficult to test independently. We briefly describe a forthcoming cluster comparison paper for which many simulators ran their codes. We were included among the testers and the code performed extremely well.

**Shock Tube**

The first of the two gas test problems is the shock tube problem of Sod (1978). The problem was run in three dimensions with initially uniform values and periodic boundary conditions in the $y$ and $z$ directions. To employ a one-dimensional version of this test is just to repeat of the early work of Monaghan and Gingold (1983). In one dimension the particles may not slip around each other to penetrate past shock boundaries or accumulate small transverse noise velocities (discussed below) and thus the test is meaningless. We advocate that all tests must be done in the number of dimensions used in the final code to give correct indications of the best parameter choices.

The initial conditions were:

\[
\begin{align*}
\rho &= 1.0, \quad P = 1.0, \quad v = 0 \quad \text{for } x < 50, \\
\rho &= 0.25, \quad P = 0.1795, \quad v = 0 \quad \text{for } x \geq 50,
\end{align*}
\]

(2.83)

with the ratio of the specific heats chosen as $\gamma = 1.4$. Values for the artificial viscosity coefficients of $\alpha = 1$ and $\beta = 2$ were used and an artificial thermal conduction coefficient of $\gamma = 0.5$.

This problem tests the variable smoothing length algorithms. The initial particle separation in the lefthand, denser region is 60% that in the right hand region. The particles all have the same mass. We started with a glass initial
condition. Particles in a realistic simulation are not (except initially) arranged on a regular grid, but pseudo-randomly in a liquid like lattice. If placed in a regular lattice and perturbed slightly the particles will, over a very long period (∼ 500 steps) commence small motions with a small associated energy conservation error and then conserve energy tightly, indicating that a regular lattice is not a minimum energy configuration (Wadsley (1993)). We generate a periodic glass initial condition by heavily perturbing the particles from an initial regular grid, providing them with pressure and a damping term on the velocity and then evolving until the velocities become small (1000 steps). This both eliminates the problem of the initial condition's instability to disorder and also tests a realistic case, without the particles colliding in rows.

The results are shown in figures 2.12, 2.13 and 2.14.

The exact solution to this problem consists of a rarefaction wave moving left, a contact discontinuity in density and thermal energy moving right and a shock moving faster to the right. These features are reproduced well, with correct shock positions and velocities at all times. This indicates that basic physics such as linear sound wave propagation is being done correctly. There are no post-shock oscillations and the shock is spread over ∼ 4h. Due to the smoothing inherent in the method, (e.g. used to calculate density and other variables), sharp discontinuities will always be spread over at least ∼ 4h in SPH calculations. It is the smoothed aspect of SPH that makes it such a robust, stable method.

Figure 2.14 uses the algorithms we have settled upon as the best for our calculations. The suppression of artificial viscosity in favour of minimizing artificial viscous transport of angular momentum is only an issue for gaseous disks and other highly dissipated systems. We do not specifically aim to model such systems and our greatest concern is to model the shocks accurately in the typical medium. We thus avoid the use of limiters and employ standard artificial viscosity which is well proven to damp unphysical post-shock oscillations. The two energy formulations discussed are shown here to be equivalent. We employ the Benz (1990) formulation
Fig. 2.12.— 3D shock tube problem, with no switch for the reduction of artificial viscosity and the older energy equation formulation, 2.17. The upper panel is at step 50 and the lower at step 100 (steps roughly constant). The solid line is the exact solution, the points are sample particles and the dots are a set of particles selected to have the mean interparticle spacing at their separation.
Fig. 2.13.— 3D shock tube problem, with a switch for the reduction of artificial viscosity and the older energy equation formulation. 2.17. Other details are identical to the preceding figure. Note the larger spread in the velocities. This is because with the viscosity limited the shock energy in oscillations (which are slightly apparent) is being not as efficiently converted into heat.
because its performance at low densities and temperatures is superior and it completely avoids any need for special modifications to the code to avoid overcooling (as earlier discussed).

\textit{Gaussian Adiabatic Wave}

A new hydrodynamical test is the propagation of adiabatic non-linear waves. The wave profile chosen for this simulation was a Gaussian. The profile chosen was:

\[
\rho(x) = 1 + e^{-0.01(x-75)^2}.
\]  

(2.84)

The initial conditions consisted of an adiabatic non-linear disturbance in the density, velocity and internal energy. To make the calculation adiabatic, internal energy must be a function of density. To have a wave travelling in only one direction, velocity must also be a function of density.

\[
\begin{align*}
    u(x) &= \frac{1}{\gamma - 1} \rho(x)^{\gamma - 1}, \\
    v_x(x) &= v_x(0) + \frac{2\sqrt{\gamma}}{\gamma - 1} \left[ \rho(x)^{\frac{\gamma+1}{2}} - \rho(x)^{\frac{\gamma-1}{2}} \right].
\end{align*}
\]

(2.85)

\(u\) and \(v_x\) refer to the initial internal energy and \(x\)-velocity respectively. This results in wave structure present in only the right propagating characteristics. The exact solution may be calculated by the method of characteristics. The wave being purely right-moving means that the left propagating characteristics all carry the same information, making solution of the characteristic equations trivial. After shock formation (when the characteristic solution becomes multivalued), the equal area method of Whitham (1974) gives an exact density solution. The position of shock formation is determined by the point on the curve that has the steepest gradient in propagation speed, \(x\)-velocity plus the sound speed, \(v_x(x) + c(x)\), at the initial time \((x = 75 + 8.267, v_x + c = 1.788)\). All points move rightward with a fixed propagation speed until the formation of the shock.
Fig. 2.14.— 3D shock tube problem with no switch for the reduction of artificial viscosity and the energy equation formulation of Benz, 2.19. Other details are identical to the preceding figures. As is visible in the figures, the energy equation choice makes no difference in this case. The Benz formulation is more accurate in the presence of large temperature variations.
Fig. 2.15.— A Gaussian wave simulated in 3 dimensions with a glass initial condition. The points are a sample of particles. The diamonds are a set of particles sampled at the mean particle spacing, indicative of the resolution. The dashed vertical line marks where the wave begins to break. The solid line is the analytical solution. Note the small amount of pre-heating occurring in the entropy plot. The artificial viscosity creates entropy as the gradient becomes very steep, anticipating the correct onset of shocking. The times shown are \( t = 0.0, 7.18, 14.7, 29.7 \) and 45.4.
We use uniform mass particles and thus the variation of density should provide a test of variable smoothing lengths in addition to the ability of the scheme to model non-linear wave propagation. An algorithm based on a mapping from a uniform-density, uniform-mass set-up is presented in appendix D. We used glass initial conditions as discussed in the previous test, compressed to produce the desired density distribution.

The simulation results are shown in figure 2.15. The results are excellent, especially the density and velocity structure. The wave-front motion has been followed extremely well. The only inaccuracy is in the entropy production. The artificial viscosity scheme detects shocks through velocity gradients and thus will anticipate the shock before the exact point of shock formation (when the velocity becomes theoretically discontinuous), as the gradients become steep, converting some kinetic energy to heat. The method resolves structure on the scale of 3-4 particle lengths, so that preheating over this distance is unavoidable. Note also a slight residual entropy associated with the initial wave due to inaccuracies in the initial condition setup.

Some upwind-type numerical schemes (e.g. PPM) employ empirical shock sharpening techniques not available to SPH. These are not always well motivated: for example, some MHD non-linear waves do not steepen with time. For a purely hydrodynamical shock they tend to give answers that look better. These schemes estimate the presence of a shock with nearby information (3 cells). Though these schemes do resolve shocks better, it is worth remembering that they rely on a few cells worth of information to do so and thus claims of resolving shocks with 1-2 cells are optimistic. It would be interesting to see the performance of these schemes on a case such as this, where the solution is not composed of constant values for physical quantities with discontinuities. Another positive aspect of SPH is its isotropic nature. High-order shock capturing schemes solve along each axial direction in order, relying on simple averaging to capture shocks travelling in non-principle directions. Spherical blast wave tests can thus demonstrate directional dependence. We respond
to the common criticism of SPH as having poor shock resolution by pointing out that inherently one-dimensional tests will show grid methods in the best light and that in higher dimensions. SPH is quite competitive.

**Self-Similar Gravitational Collapse**

Following Couchman *et al.* (1995), we performed a self-similar gravitational collapse of a top hat density perturbation. We chose the initial perturbation at $z = 60$ so that the linearly evolved overdensity at $z = 0$ would be $\delta = 1.686$. Results are shown in figures 2.16 and 2.17. We began with a glass initial condition determined for the previous tests, as we did not want any artifacts associated with grids in the results. The total number of particles used was $32^3/2$.

The analytical expression for the collapse of a uniform top hat density perturbation is given parametrically (Peebles 1971).

\[
\begin{align*}
    r &= A(1 - \cos \theta). \\
    t &= B(\theta - \sin \theta). \\
    A^3 &= G M B^2.
\end{align*}
\]

(2.86) \hspace{1cm} (2.87) \hspace{1cm} (2.88)

where $M$ is the mass enclosed by the radius $r$. If we assume $\theta$ and thus $t$ is small, we can expand to find a relationship between linear overdensity $\delta$ and $B$.

\[
\begin{align*}
    1 + \delta &= \frac{\rho}{\bar{\rho}} \propto r^{-1} t^{-2}. \\
    1 + \delta &= 1 + \frac{3}{20} \left( \frac{6t}{B} \right)^{2/3}, \\
    B &= 6 \left( \frac{3}{206} \right)^{3/2} t_0.
\end{align*}
\]

(2.89) \hspace{1cm} (2.90)

$t_0$ is the current age of the universe, at $z = 0$. The collapsing object turns around at $\theta = \pi, z = 0.59$ and becomes singular at $\theta = 2\pi, z = 0.0$. The singular part of the collapse is very difficult to model correctly, made more difficult by the noise present in the initial condition. A glass formed using gravity would probably produce a less
noisy result, but we are satisfied that the current result demonstrates the ability to follow the collapse accurately. The final overdensity is enormous, \( \delta \sim 4000 \). Singular collapse is an extreme case that does not occur in reality because there is always substructure. The collapsing cloud is unstable to the formation of substructure. The glass initial conditions have density variations around the 1% level. This is comparable to the density perturbation added in the initial condition. However, they do not have the accompanying velocity perturbations which are more important for the evolution. If this were not the case, the whole top hat would have fragmented, forming substructure.

The parameter \( B \) is the same for infalling shells at all radii. \( \mathcal{M} \) is fixed at \( 4\pi/3\rho_{\text{init}}r_{\text{init}}^3 \) for each shell as the shells do not cross before \( \theta = 2\pi \). The collapse is thus self-similar and \( v = dr/dt = rf(t) \), where \( f \) contains at the time dependent information. The linear relationship between infall velocity and radius is clearly seen in figure 2.16.
Fig. 2.16.— Top hat collapse. The perturbation linearly evolved to $z=0$ would have $\delta = 1.686$. The redshifts shown, in order of increasing collapse, are $z = 1.0$, $0.3$, $0.1$ and $0.0$. In the left panel, we plot density times $a^3 = (1 + z)^{-3}$ so the mean density is fixed at the right hand edge. In the right panel we show radial infall velocity, which the self-similarity solution dictates must be a linear function of radius.
Fig. 2.17.— Top hat collapse. The perturbation linearly evolved to $z=0$ would have $\delta = 1.686$. The redshifts shown (in order of increasing collapse) are $z = 2.0, 1.0, 0.5, 0.3, 0.1$ and $0.0$. We plot density times $a^3 = (1 + z)^{-3}$ so the mean density is constant in the figure. The maximal expansion is clearly evident at $z = 0.59$. A singular state is approached at $z = 0.0$. The lines mark the analytical solution.
Cluster Comparison test

At the Institute for Theoretical Physics meeting on Galaxy Evolution in 1995, it was suggested that a cluster initial condition be simulated by many computational methods to both validate and compare schemes and to identify the features of the result that are independent of the method used. The full details of the cluster comparison will be reported in Frenk et al. (1998). The cluster run was a test of both the 3-dimensional gravity hydrodynamics code and of our approach of using high resolution volumes surrounded with tidal low resolution particles with external tides applied. The comparison focussed on the inner part of the cluster, so were able to simulate the outer parts at lowered resolution without compromising the comparison.

The cluster initial condition was generated at a resolution of $256^3$ covering a 64 Mpc comoving box. The cluster perturbation corresponds to a $3\sigma$ peak for a Gaussian filter with radius 10 Mpc. The simulations were all adiabatic (no atomic heating or cooling processes) and commenced at redshift $z = 20$. Statistics were measured on the inner 16 Mpc comoving radius sphere and in a spherical volume with a mean overdensity of 200, centred on the centre of mass.

The Gaussian filter scale for the peak was too large for the periodic box size to adequately treat the tidal environment (i.e., the cluster was apodized). Normally we would add tidal particles out to at least 2.5 times the high resolution region. To our spherical high resolution region of radius 25 Mpc, we could only add a single lower resolution shell extending to 32 Mpc in radius. The 25 Mpc choice was based on the region the peak patch theory (Bond and Myers 1996) suggests would have collapsed. The low resolution particles had 8 times the mass of the high resolution ones. A self-consistent linearly-evolved external shear was also applied to the entire region. There were 74127 gas and 74127 dark particles used in the simulation. The initial $256^3$ displacement field was sampled at every fourth lattice site to transfer onto the computational grid for the high resolution region. A similar transfer was done for
the low resolution region, with slight smoothing added. Couchman used the same one-in-four transfer method and his results are visually similar. Discrepancies may be due to his use of periodic boundary conditions.

For the paper, the SPH-P³MultiGrid code was used. The computation was run on a Dec-Alpha EV5 and required 119 CPU Hours and 100 Mb of memory. With Tree-pp and improved neighbour finding, the same computation takes 33.6 CPU Hours and 50 Mb of memory. The results are identical. Timing comparisons for specific parts of the code are given in preceding sections.

A numerical comparison between code results is shown in figure 2.18. This is a preliminary figure from Frenk et al. (1998) and some numbers must be modified for the final version. In particular for our result, \( \frac{U_{\text{kin}}}{U_{\text{therm}}} = 0.10 \). The change is required because we measured our kinetic energy in the centre of mass frame initially, instead of the simulation rest frame. This also moves the \( J \) measure up to roughly 1.11. Our simulation is more dynamically advanced than most other calculations at \( z=0 \), and is thus hotter and more relaxed.

5. Time steps and Integration

The simulations were stepped forward in time with a Predictor-Corrector scheme. The time steps are variable and set at the end of each step. The calculation commences with a time step of zero, to evaluate all the forces. Calculating the first non-zero time step requires information generated in a computational step.

The forces from the previous step are used to predict the velocities for the first half of the step. These velocities predict the midpoint positions. The predicted velocities are then updated to the midpoint predicted values using the old forces. The new forces are calculated at the midpoints using the predicted midpoint velocities and positions. The method is second order.

The time step is limited by two standard conditions, the amplitude of the
Fig. 2.18.—Cluster Comparison. The quantities are measured within the virial radius of the cluster. On the left, from top to bottom are: cluster mass, one dimensional dark matter velocity dispersion, gas mass fraction and axial ratios of dark matter (solid) and gas (open circles). On the right from top to bottom are: mass-weighted gas temperature, $\beta = \mu m_p \sigma_{DM}^2 / kT$, the ratio of bulk kinetic to bulk thermal energy in the gas and the X-ray luminosity.
acceleration and the Courant condition. The acceleration condition is calculated so that the time step is given by,

$$\Delta t_{\text{acc}} \leq \min_{\text{particles}} 0.2 \sqrt{\frac{h}{|a_{\text{particle}}|}}.$$  

(2.91)

where $a_{\text{particle}}$ is the acceleration of the particle and $h$ is the minimum of the gravity softening and the SPH gas particle softening. For our calculations they are the same and constant. This condition limits the rate of change of the velocity.

$$\Delta t_{\text{Courant}} \leq \min_{\text{gas pairs}} 0.4 \sqrt{\frac{h_{\text{pair}}}{c_{\text{pair}} + \zeta_{\text{pair}}}}.$$  

(2.92)

where $c_{\text{pair}}$, $h_{\text{pair}}$ are the average sound speed and smoothing and $\Delta v_{\text{pair}}$ and $\Delta r_{\text{pair}}$ are the velocity difference and separation of an interacting gas particle pair. The $\zeta_{\text{pair}}$ term is only added for approaching pairs (these are pairs possibly undergoing shocking).

In practice it is the Courant condition which is the more important one. This has led some authors to implement variable time steps for gas particles only (Hernquist and Katz (1989)). We investigated this possibility and found that it would require a large coding and testing effort.

6. Evolution of the Lyman-\(\alpha\) runs

While the code has undergone several improvements, there have also been substantial changes in the numerical parameters of the simulations, motivated by accuracy and comparisons with observations.

Initial Lyman-\(\alpha\) runs used $32^3$ gas and $32^3$ dark matter particles with a shell cut from $32^3$ dark matter and $32^3$ gas low resolution (twice the spacing) particles to act as a buffer around the sharp edge to the high resolution inner region. The shells
had radii of 3.2 and 6.4 Mpc comoving respectively for 0.1 Mpc particle spacing.

We realised that we had the computing power to do larger simulations in a reasonable time and expanded to $50^3$ high resolution particles of each type. We first used the $32^3$ low resolution buffer but later expanded it to $40^3$ to ensure that even highly collapsed calculations would have a buffer. We also added a further shell cut from $32^3$ dark matter and gas combined collisionless particles with four times the spacing of the high resolution to model the tides better. Our final production code thus operated with high resolution to 5 Mpc, half resolution to 8 Mpc and lowest resolution collisional particles out to 12.8 Mpc.

The initial runs used a temperature floor of 8000 K. Partial motivation for this was to prevent the time step getting too small by maintaining sufficient pressure to inhibit extreme collapses to very high densities in objects. This was not an unphysical thing to do. because processes such as star formation would be expected to maintain the temperature. It turned out that at the densities typical of collapsed objects forming in the simulations, this was not a problem and the equilibrium temperatures remained high. We lowered the floor to 1000 K because low density gas can reach this temperature. This change had a very limited effect on the Lyman-α forest parameters derived from the simulations. All the results presented in Chapter Four were simulated with the new temperature floor.

7. Convergence

Convergence is a major issue confronting simulators. We must be able to show the results will not change if go to still higher resolution. We performed two limited studies of convergence. First, we performed a standard simulation with half the initial spacing, 8 times worse resolution in mass. We also performed a complete set of simulations with initial spatial resolution 0.3125 comoving Mpc. This set of simulations resolve larger dwarf galaxies and was intended for that purpose. All the lowered resolutions simulations produce less absorption lines compared to high
Fig. 2.19.— A sample los through the $\nu = 0$ CDM simulation at $z = 3$. The panels compare the a standard resolution run (100 kpc comoving initial grid) with a run with twice poorer resolution (same initial condition). The structures are quite changed due to the additional small scale structure resolved. The mean opacity is clearly lowered at lower resolution. The grey lines are Voigt fits.
Fig. 2.20.—A sample los through the $v = 0$ CDM simulation at $z = 3$. These two figure compares the same line of sight through simulations with different resolution. See figure 4.8. This clearly shows how insufficient resolution is incapable of correctly resolving the filaments that give rise to Lyman-$\alpha$ absorption.
resolution when all other parameters are held constant.

In figures 2.19 and 2.20 we show artificial spectra generated from high and medium resolution simulations with the same initial conditions.

In figure 2.21 we compare the results for the distribution of line column densities at the three resolutions listed. The ultraviolet-flux level was $J_{-21} = 0.5$ for the artificial spectra generation for all cases. This flux is slightly high but reasonable. The poor result for the lower resolution cases demonstrates that our resolution should be considered the minimum for IGM simulations that intend to model the Lyman-$\alpha$ forest well. None of these cases represents a complete simulation set, but the cases shown are fairly representative. They tend to be slightly higher than a combined result.

In all the cases but one the initial condition resolution shown in the figure is the one used for the gas during simulation. The curves labelled “HR $l_{\text{mass}} = 0.2$ Mpc” represent a high resolution run with initial conditions smoothed to the initial grid spacing of the “MR” medium resolution run. Between the two high numerical resolution runs, the additional structures associated with the greater mass resolution do not have much impact on the Lyman-$\alpha$ forest column density distribution. The break-up of the filaments due to the additional small scale waves results in fewer low column lines ($N_{HI} \sim 10^{13} \text{ cm}^{-2}$).

In figure 2.23 we show the visual effect that changing the resolution has on the gas columns. Comparing the upper left and lower left panels we see that smoothing the initial conditions does not affect the major filaments directly, but removes the short waves that promote their break up. Many of the the major collapsed structures have still formed and some of the smaller ones. This also demonstrates that the code successfully evolves waves down to the smallest grid scale (less than two particle spacings). A particle-particle gravity solver was essential to do this.

The lower right panel demonstrates the greater effect of reducing the gas resolution as well. This simulation had twice the initial grid spacing. The results are
Fig. 2.21.— Column density distribution for three resolutions at $z = 3$. The resolutions are specified using initial comoving grid spacing. The masses of the particles are given by this spacing cubed. The medium resolution misses the small objects that are responsible for most of the absorption lines. Each curve represents a single simulation with $\nu = 0$. The medium resolution has the same initial condition as the high resolution with 1/8th as many particles. The low resolution sample is the $\nu = 0$ galaxy scale simulation with 1/30th as many particles per unit volume. The last curve, labelled 'HR $l_{\text{mass}} = 0.2 \text{ Mpc}$' is a high resolution run with smoothed initial conditions.
Fig. 2.22.— Frequency of lines of a given column density $N_{HI}$ at $z = 3$ for simulations with the same initial conditions and different mass resolutions (marked according to the comoving initial grid sizes, $l_{mass}$). The simulations used the same ionizing flux normalization. The errors bars are Poisson and the deviation from the observed data is expected, due to the small box size. A significant excess of higher column lines occurs in the lowest resolution run, due to the larger size of condensed objects.
Fig. 2.23.— Column density through 2 Mpc thick slices at $z = 3$. These panels show the effect of different resolution. The initial conditions were the same for all panels ($\nu = 0$, SCDM) and the flux level used for the plotting was $J_{-21} = 0.22$. The upper panels are the same: the standard simulation. The lower two panels have initial conditions smoothed to twice the spacing. The lower right panel was run with 1/8th as many particles (appropriate to the smoothing). The contours in $\log_{10} N_{HI}$ are 11-12.75 (dark), 13 (dark grey), 13.5 (medium grey), 14 (light grey), 14.5 (dark), 15 (black) and 17 (white).
very different, with filaments largely absent. This is what produced the drastic fall in the line numbers shown in figure 2.22.

To demonstrate that our resolution is not only an upper bound but sufficient to give an accurate results for Lyman-\(\alpha\) forest absorbers, we performed simulations with a smaller box - too small to provide a reasonable sample of the universe. We ran the simulation at 3 resolutions: 0.1, 0.05 and 0.025 Mpc. The largest spacing is the same as our standard initial spacing of 0.1 Mpc.

The classic Jeans scale is given in the Chapter One. It corresponds to a mass that evolves with redshift, of order \(10^9(4/(1+z))^{3/2}h^{-1}\Omega_{nr}^{-1/2}M_\odot\). \(\Omega_{nr}\) is the density in non-relativistic matter (dark matter and baryons). For our standard high resolution the mass of our dark matter particles is \(2.78\times10^6\Omega_{dm}h^2M_\odot\). The gas particles have masses \(3.48\times10^6h^{-1}M_\odot\) for every cosmology (by design). Pressure represents the only limiter on the formation of dense gaseous lumps. The mean dark matter fluctuation amplitude only gets larger as we move to smaller scale objects. The presence of dark matter potential wells has the effect of lowering the effective Jeans mass relative to a purely gaseous medium. If the dark matter collapses, this effect becomes greater because it is actually the ratio of dark matter to gas that is important (expressed via \(\Omega_{nr}r^{-1/2}\) in the equation) and this ratio increases if the gas does not collapse similarly.

Our convergence studies indicate that our resolution is sufficient to resolve this lowered scale - we do not find that more dense gas clumps are formed as we increase the mass resolution to more than 1 particle per \((100\text{ kpc})^3\) comoving. The structure of the objects does change. When objects are simulated by only a few particles the degree of collapse is limited. This makes the cores of these objects larger as seen in figure 2.24 in the top panel. This effect also occurs partly because the high resolution initial conditions had to be smoothed to the low resolution initial spacing. This affects primarily lines above \(N_{HI} = 10^{15}\text{ cm}^{-2}\), as shown in figure 2.22. The lines in the figure were Voigt profile fit in artificial spectra generated as for previous simulation output. More resolution is always desirable, but the single filament
Fig. 2.24.— Column depth through a cube 800 comoving kpc thick at $z=3$. From top to bottom the resolution increases with initial grid spacings of 100, 50 and 25 kpc respectively. The left panels show integrals along the line of sight of Dark Matter density squared and the right hand panels show neutral hydrogen ($\propto$ gas density squared). The contour levels for $N_{HI}$, in $\log_{10}(\text{cm}^{-2})$, are 12, 12.25, 12.5, 12.75 (dotted), 13 (solid), 13.36 (mean density, dot dash) 14 (solid), 14.5 (dashed), 15 and 17 (solid).
shown in the figure is hardly a sufficient sample of the universe. This prohibits using smaller simulations to achieve better resolution. Computing power is always improving so future work on this subject should employ better resolution.

The additional physics of radiative transfer, star-formation and supernova heat injection should be added to model collapsed objects correctly. These may act to puff out the cores of these objects in a similar fashion to numerical effects associated with resolution. Star formation in dwarf galaxies is the subject of current research (e.g. Klypin et al. (1998)).
Appendices

A. Estimating the physical analogue of SPH expressions

In one dimension it is fairly straightforward to show that the SPH equations approach a given continuum limit. Three equispaced particles readily yield an appropriate expression in general. One can attempt this in two or higher dimensions, but the laying down of regular grids imparts a level of directionality to the estimate.

Integrating over a continuous ring of material provides a non-directional estimate in two dimensions as follows:

The particle masses \( m_j \) in the SPH equation are replaced by \( A \rho r d\theta \), where \( A \) is a constant included only to maintain correct scaling. The sum is then converted to an integral over \( \theta \).

For example, the linear term in the standard expression for artificial viscosity at particle \( i \) (Equation 2.23):

\[
\sum_j \alpha \frac{\bar{c}_{ij}}{\bar{\rho}_{ij}} \left( \frac{h(v_{ij} \cdot r_{ij})}{r_{ij}^2} \right) \nabla_i W_{ij},
\]

is represented as:

\[
\int_0^{2\pi} \frac{\alpha c h}{r^2} (v_j - v_i) \cdot \mathbf{r} (\nabla W) A r d\theta, \tag{2.94}
\]

where \( \rho, c \) and \( h \) have been assumed constant and \( \mathbf{r} = r_j - r_i \). The following substitutions are made:

\[
\begin{align*}
\mathbf{r} & = x \hat{e}_x + y \hat{e}_y, \\
\mathbf{v}_i & = \mathbf{v} \\
\mathbf{v}_j & = \mathbf{v} + (\mathbf{r} \cdot \nabla) \mathbf{v} + (\mathbf{r} \cdot \nabla)^2 \mathbf{v} + O(r^3) \\
x & = r \cos \theta
\end{align*}
\]
and

\[
y = r \sin \theta
\]  

(2.95)

In this case, the expression reduces to:

\[
\alpha c h A r^2 \left( \nabla W \right) \frac{\pi}{8} \left[ \nabla^2 \mathbf{v} + 2 \nabla(\nabla \cdot \mathbf{v}) \right] + O(r).
\]  

(2.96)

This would indicate that in the limit as \( r \) becomes small (with \( h \sim r \) and \( \nabla W \sim r^{-3} \)), the linear artificial viscosity term in two dimensions is approximating:

\[
\eta \nabla^2 \mathbf{v} + \left( \frac{\eta}{3} + \zeta \right) \nabla(\nabla \cdot \mathbf{v}).
\]  

(2.97)

with \( \zeta = \frac{5}{3} \eta \). This is the Navier-Stokes form for viscosity, taken from Landau and Lifshitz (1987), with \( \eta \) the coefficient of shear viscosity and \( \zeta \) the coefficient of bulk viscosity.

B. Anisotropic kernels and Tensor \( h \)

Some phenomena have a preferred direction or large differences in the length scales in different directions. These include disks and Zel’dovich pancakes. In these situations, SPH with spherical kernels suffers because it is has inadequate resolution in one or more directions. For example, consider a two dimensional SPH model of a velocity field such as \( \mathbf{v} = (-ax, 0) \). The size of the kernel will shrink in all directions, whereas the compression is only occurring in the x direction. An anisotropic kernel that varies its shape in time is the natural answer. However, such kernels produce spurious angular momentum transfer because the force between a particle pair will not in general occur along the vector connecting the two particles, as described in 7.. Shapiro and Martel (1993) have developed a version of SPH with anisotropic kernels. These kernels have no intrinsic preferred direction. Thus six numbers are required to fully specify the shape of the kernel in three dimensions, as can be
represented by a symmetric tensor $H_{\alpha\beta}$. The range of influence of the kernel is described by an ellipsoid:

$$x^T H x = H_{11} x^2 + 2H_{12} xy + \ldots + H_{33} z^2 = \text{constant}. \quad (2.98)$$

Shapiro and Martel evolve $H$ using the Strain Tensor, so that to first order the shape of the kernel changes in the same way a small fluid element of the same shape would. They calculate the primary, secondary and tertiary axes of this ellipsoid, and their directions and construct a kernel of the form

$$W(r, H) = W_1(l_1, h_1)W_2(l_2, h_2)W_3(l_3, h_3),$$

where each $W_i$ is a Gaussian. It is not clear that the authors avoid the problem of spurious angular momentum transfer.

**B.1 Implementing Tensor $H$**

An ellipsoid described by the following function equal to a constant:

$$f = H_{\alpha\beta} x^\alpha x^\beta$$

$$= H_{11} x^2 + 2H_{12} xy + 2H_{13} xz + H_{22} y^2 + 2H_{23} yz + H_{33} z^2. \quad (2.99)$$

can be made to evolve and deform with the fluid if $f$ satisfies:

$$\frac{df}{dt} = \frac{\partial f}{\partial t} + (\mathbf{v} \cdot \nabla) f = 0 \quad (2.100)$$

If $\mathbf{v}$ is expanded to first order only around the origin, $x = (x, y, z) = 0$, then the ellipsoid will be deforming in the same way as an infinitesimal fluid element at the origin, or equivalently, the deformation of the ellipse is determined by the strain tensor at origin. Thus we solve:

$$\frac{\partial f}{\partial t} + ((\mathbf{v}_0 + (x \cdot \nabla)\mathbf{v}_0) \cdot \nabla) f = 0. \quad (2.101)$$
for the coefficients $H_{\alpha\beta}$ for all $x,y,z$. The resulting expressions are:

\[
\begin{align*}
\frac{dH_{11}}{dt} &= -2H_{11} \frac{\partial v^1}{\partial x^1} - 2H_{12} \frac{\partial v^2}{\partial x^1} - 2H_{13} \frac{\partial v^3}{\partial x^1} \\
\frac{dH_{12}}{dt} &= -H_{12} \left( \frac{\partial v^1}{\partial x^1} + \frac{\partial v^2}{\partial x^2} \right) - H_{11} \frac{\partial v^1}{\partial x^2} - H_{22} \frac{\partial v^2}{\partial x^1} - H_{13} \frac{\partial v^3}{\partial x^1} - H_{23} \frac{\partial v^3}{\partial x^2}
\end{align*}
\]

and so on...

The kernel can be rewritten by replacing $r_{ij}^2/h^2$ by $f$ in the formula. The $H_{\alpha\beta}$ are like $1/h^2$. The normalisation of the kernel is more complex. Given:

\[
\int W^*(x^2)d^3x = 1
\]

\[
\int \sqrt{\det(H_{\alpha\beta})} W^*(H_{\alpha\beta}x^\alpha x^\beta)d^3x = 1.
\]

The gradient of $W$ is written (ignoring terms due to spatial variation of $H$):

\[
\begin{align*}
W(r - x, H) &= W(f) \text{ where } f = (r - x)^TH(r - x) \\
\nabla_r W(r - x, H) &= \frac{\partial W}{\partial r^\alpha} \hat{e}_\alpha \\
&= \frac{\partial f}{\partial r^\alpha} \frac{dW(f)}{df} \hat{e}_\alpha \\
&= (2H_{11}(r^1 - x^1) + 2H_{12}(r^2 - x^2) + 2H_{13}(r^3 - x^3)) \\
&\quad \frac{dW(f)}{df} \hat{e}_1 + \ldots \\
&\neq (r - x).
\end{align*}
\]

Whereas for a spherical kernel (ignoring terms due to spatial variation of $h$):

\[
\begin{align*}
W(r - x, h) &= W(u) \text{ where } u = \frac{(r - x)^2}{h^2} \\
\nabla_r W(r - x, h) &= \frac{\partial W}{\partial r^\alpha} \hat{e}_\alpha \\
&= 2(r - x) \frac{dW(u)}{du}
\end{align*}
\]
Herein lies the spurious angular momentum transfer. Unlike the case for spherical kernels, $\nabla_i W_{ij}$ does not in general point along the centreline between a particle pair and thus angular momentum will not be conserved.

C. Modification to standard SPH Kernels

The leading $O(h^2)$ error term of equation 2.5 that related the smoothed estimate to the local value can be eliminated at the expense of allowing the kernel to be negative for some values of $r$. Such kernels may be derived by modifying existing kernels according to:

$$W_{new}(u) = BW_{old}(u)(1 - Au^2)$$  \hspace{1cm} (2.107)

and solving for $A$ and $B$ such that the new kernel satisfies equation 2.2 and the second moments vanish (Monaghan (1992)). Such kernels have leading error terms $O(h^4)$. Benz (1990) argued against the use of such kernels because the negative values involved would imply a negative mass being used for gravitational calculations (from a scatter interpretation of the interpolation equations. It is also possible that all the negative contributions from surrounding particles may be sufficient to drive the value of a positive definite quantity such as density negative in some region. It has been shown (Monaghan and Gingold (1983), Lattanzio et al. (1986)) that judicious use of the Super-Gaussian kernel (an $O(h^4)$ version of the Gaussian kernel with a negative part) makes for good sharp shocks in SPH simulations. However, they did experience problems due the non-monotonicity of the Super-Gaussian: leading to a spurious bump in the density and velocity profiles on either side of the shock and to instability in high Mach number ($\sim 10$) calculations. Lattanzio et al. (1986) tried a combination of a Gaussian and a Super-Gaussian, such that the kernel was Gaussian near shocks, with good results.

A common property of the kernels mentioned above is that the kernels are flat at $r = 0$. The force term is proportional to the derivative with respect to $r$ of the
kernel, which results in the force becoming small when particles are within \( \frac{1}{2} h \) or so. The contribution of particles outside the close pair can lead to the two particles clumping together. This is often called the pairing instability. Shüssler and Schmitt (1981) suggested that centrally peaked kernels could solve this problem. They offered:

\[ W(u) \propto (1 - \sqrt{u}), \quad \text{for} \quad 0 \leq u = \frac{r}{h} \leq 1, \quad (2.108) \]

as one alternative. For this kernel the force increases monotonically as the particles come closer together. It has poor interpolating qualities, however.

The problem might be tackled through the use of a centrally peaked kernel for the force calculation and a good interpolating kernel for the density calculation. Conservation of energy and momentum require only that the same kernel be used for energy and momentum equations and that particle-particle interactions be symmetric respectively. In practice, particle clumping is rarely problematic and we do not see it in our calculations, even though they are of a fairly dynamic nature.

**D. Density profiles with constant particle mass in 1D**

Given two coordinate systems \( x \) and \( \xi \), one in which the density, \( \rho(x) \), is unity, and one in which the density has the desired form \( \rho(\xi) \), the fact that \( \rho(x)dx = \rho(\xi)d\xi \) provides the equation:

\[ \frac{d\xi}{dx} = \frac{1}{\rho(\xi)}, \quad (2.109) \]

which may be integrated to provide the necessary transformation.

Say that the profile \( \rho(r) \) is to be reproduced for use in a one-dimensional code. All particles have equal mass. Define \( \Delta x \) as the particle separation that will result in a density of 1.
Given particle \( i \) at \( r = r_i \), then the next particle should be placed at:

\[
r_{i+1} = \int_0^{\Delta x} \frac{dr}{dx} \, dx,
\]

where the \( r(x = 0) = r_i \) and \( dr/dx = 1/\rho(r) \) as defined by 2.109. This integration can be performed with a scheme such as Runge-Kutta with \( \Delta x \) as the step size.

### E. Two-Dimensional Gravity

The earliest version of the full multigrid gravity solver was developed in two-dimensions; suitable for simulations of rotating disc, stars or gas clouds or pancake collapses.

Poisson’s Equation in two dimensions is written:

\[
\frac{\partial^2 \phi(r)}{\partial x^2} + \frac{\partial^2 \phi(r)}{\partial y^2} = 4\pi G \rho(r).
\]

A point source, \( \rho(r) = M \delta(r - r_0) \), in two dimensions does not produce the same potential as a point source in three dimensions. The three-dimensional equivalent is an infinite line of mass extending perpendicular to the two-dimensional plane. As a result, the force and the potential, which would be proportional to \( \frac{1}{r - r'} \) and \( \frac{1}{r - r'} \), respectively, in three dimensions, are:

\[
\text{Force} = -\frac{2GM}{|r - r'|^2} (r - r'),
\]

\[
\text{Potential} = 2GM \log_e |r - r'|.
\]

A general expression for the potential of a cylindrically symmetric source is calculated as follows:

\[
\phi(r) = 4\pi G \left[ \log_e r \int_0^r \rho(r') r' \, dr' + \int_r^\infty \rho(r') r' \log_e r' \, dr' \right].
\]

For solution of these equations using multigrid techniques, we require boundary
values and thus a multipole expansion solution to the two-dimensional Poisson’s Equation is required. An expression for the potential of a collection of point sources, with coordinates \( \{(r_\alpha, \theta_\alpha)\} \), and masses \( \{M_\alpha\} \) is as follows:

\[
\phi(r, \theta) = 2G \left\{ -\sum_{m=1}^{\infty} \frac{1}{m} \left( e^{im\theta} \left[ r^m \sum_{r<r_\alpha} M_\alpha \frac{1}{r_\alpha^m} e^{-im\theta_\alpha} + \frac{1}{r^m} \sum_{r>r_\alpha} M_\alpha r_\alpha^m e^{-im\theta_\alpha} \right] \\
+ e^{-im\theta} \left[ r^m \sum_{r<r_\alpha} M_\alpha \frac{1}{r_\alpha^m} e^{im\theta_\alpha} + \frac{1}{r^m} \sum_{r>r_\alpha} M_\alpha r_\alpha^m e^{im\theta_\alpha} \right] \right) \\
+ \sum_{r<r_\alpha} M_\alpha \log_e r_\alpha + \log_e r \sum_{r>r_\alpha} M_\alpha \right\}. \tag{2.114}
\]

Some simulations employ periodic boundaries in one direction. Such an implementation would be suitable for the numerical study of sections of discs or slabs of material. For this case the potential at the non-periodic boundary is required if the gravity is to be solved with a grid-based method. We have derived an expression for the multipole expansion in this case, but it has not been implemented. The test problem in mind was simulations of cosmological pancake collapse and fragmentation. Top down models are currently in disfavour and two dimensional codes are not capable of reproducing the full dynamics of many astrophysical problems. We thus focused on upgrading to a a full three-dimensional code.
Abel, T. 1996, Ph.D. Thesis
Benz, W. 1988, Computer Physics Communications, 48, 97
Benz, W. 1993, in Smooth Particle Hydrodynamics in Astrophysics, OAT-SISSA International Workshop, Trieste. Italy
Godunov, S.K. 1959, Mat. Sb., 47. 357
Greengard, L. & Rokhlin, V. 1986, A Fast Algorithm for Particle Simulations, research report YALEU/DCS/RR-459
Herant, M. 1993, in Smooth Particle Hydrodynamics in Astrophysics, OAT-SISSA International Workshop, Trieste, Italy

Hockney, R.W. & Eastwood, J.W. 1988, Computer Simulation using Particles. IOP Publishing


Inutsuka, S. 1993, in Smooth Particle Hydrodynamics in Astrophysics. OAT-SISSA International Workshop, Trieste, Italy


Monaghan, J.J. 1985, Computer Physics Reports, 3, 71

Monaghan, J.J. 1988, Computer Physics Communications, 48, 89


Monaghan, J.J. 1989, Journal of Computational Physics, 82, 1

Nelson. A.H. 1993. in Smooth Particle Hydrodynamics in Astrophysics. OAT-SISSA International Workshop, Trieste, Italy
Shapiro, P.R. & Martel, H. 1993, in Smooth Particle Hydrodynamics in Astrophysics, OAT-SISSA International Workshop, Trieste. Italy
Schoenberg, I.J. 1973, Cardinal Spline Interpolation, SIAM, Philadelphia
Sod, G.A. 1978, Journal of Computational Physics. 27. 1


Steinmetz. M. 1996, to be published in The Early Universe with the VLT. ESO Workshop


Chapter 3

Design

In this chapter we describe how we design initial conditions for our simulations of the Lyman-\(\alpha\) Forest. We discuss the general application of constrained-field initial conditions and present simulations targeting specific objects and structures.

Firstly, we introduce Gaussian random fields, for they describe the random density fluctuations that gave rise to present day structure in the universe and necessarily make up the initial conditions for all cosmological simulations.

Secondly, we describe the cosmological models we study.

In section 3., we motivate our use of constrained initial conditions and non-periodic simulations. We discuss our technique in the light of the power spectrum of density fluctuations and the dynamical range of numerical simulations. We discuss how current cosmological simulations fare at resolving the important scales: large and small.

In section 4. we define the important spatial variables whose values, smoothed on the scale of a small patch of the universe, determine its subsequent evolution. Following on with section 5., we summarize the probability distribution of the most important parameter: the shear tensor. Shear decomposes into the background fluctuation level \(\nu\), ellipticity \(e\), and prolateness \(p\). We constrain the shear in our
initial conditions, lending the name shearing patches to our simulation volumes.

We describe the parameter choices for our Lyman-α initial conditions in section 6.

We briefly outline the peak patch algorithm in section 7. In section 8, we give an example of a direct application of peak patches to the reconstruction of initial conditions for an interesting region located in a large simulation, resampled at high resolution.

The historical development of structure formation, understood in terms of smoothed properties of the underlying initial conditions, leads to the peak patch picture and ultimately to the cosmic web description of structure (section 9). The Cosmic Web picture provides a powerful language for understanding the structure and evolution of Lyman-α absorbers.

Finally, we describe new work, which applies the peak patch technique to study Lyman-α absorbing systems on large scales.

1. Gaussian Random Fields

The Gaussian random field description of density perturbations provides an analytical understanding of the statistical properties of the universe. A Gaussian random field, $F(\mathbf{r})$, is one which may be fully described through knowledge of the correlation function, $\xi(r_1, r_2)$. The field is homogeneous if spatial translation does not change its properties and thus $\xi$ is a function only of $r_1 - r_2$. If the field is isotropic then $\xi$ is a function of $|r_1 - r_2|$ and the Fourier transform of $\xi$; the power spectrum, $P(k) = \langle |F_k|^2 \rangle$ is a function of $|k|$ only. $F(\mathbf{r})$ may be written as

$$F(\mathbf{r}) = \sum |F_k| \cos(k \cdot \mathbf{r} + \theta_k).$$

Strictly, the modes must be independent and the phases random, with the moduli Rayleigh distributed, so that the distribution of values for each $|F_k|$ is a Gaussian
with zero mean and standard deviation $\sqrt{P(k)}$. The central limit theorem may result in non-Gaussian random fields tending towards Gaussian. For cosmology, we use this as a description of density perturbations in the early universe. The overdensity perturbation is defined.

$$\delta(r) = (\rho(r) - <\rho>)/<\rho>.$$ \hfill (3.2)

We characterize the perturbations on a given scale with the use of smooth filters such as a Gaussian.

$$W_G = \frac{1}{(2\pi)^3/2R_f^3} \exp \left( -\frac{r^2}{R_f^2} \right).$$ \hfill (3.3)

or a top hat,

$$W_{TH} = \frac{1}{(4\pi/3)R_f^3} \begin{cases} 1, & r < R_f, \\ 0, & r \geq R_f \end{cases}.$$ \hfill (3.4)

The density perturbation field convolved with these filters remains Gaussian-distributed. We can determine the rms fluctuation in $\delta$, filtered on a given scale through an integral of the power spectrum.

$$\sigma(R_f)^2 \equiv \langle \delta^2_{R_f} \rangle = \int \bar{W}^2(k)P(k) \frac{d^3k}{(2\pi)^3}.$$ \hfill (3.5)

where we have used the bar to denote the Fourier transform and $W$ may be any filter. $\sigma_8$, the rms $\delta$ top-hat filtered on a scale of $8\ h^{-1}\text{Mpc}$, characterizes the abundance of clusters and serves as a data point constraining the normalization of the power spectrum. We prefer to use a Gaussian filter because it does not ring in $k$-space.

2. Cosmologies

On the basis of the current data it seems that a correct description of the observable universe is not too far removed from the current Cold Dark matter
related models – in particular with regard to a bottom-up structure formation hierarchy. The basic problem with standard cold dark matter is that it cannot simultaneously satisfy COBE and cluster constraints. We need to suppress the power at cluster scales and smaller. There are many extensions of the basic flat, pure cold dark matter model which allow this (Bardeen, Bond & Efstathiou (1987)). Options include a tilt to the scale invariant initial spectrum derived from inflation, abandoning flatness for open models, including hot dark matter and models with vacuum energy.

We have simulated four cosmologies: standard cold dark matter, open cold dark matter, vacuum energy ($\Lambda$) plus cold dark matter and a mixed hot plus cold dark matter model. These choices are heavily constrained by the Cosmic Microwave Background anisotropies and clusters abundances. We ran three normalizations for the standard CDM case. The models all have $\Omega_b h^2 = 0.0125$, $\sigma_0.5 = 1.05$ at redshift $z = 3$. $\Omega_b$ refers to the fraction of the critical mass in baryons at $z = 0$, so that we have the same mass of baryons in our simulation volume for each cosmology. The Hubble parameter is given as $H_0 = h \times 100 \, \text{km s}^{-1} \, \text{Mpc}^{-1}$. $\sigma_0.5$ refers to the rms Gaussian smoothed fluctuation level in the density with a filter scale of $0.5 \, \text{Mpc}$. As discussed in Chapter One, this scale is characteristic of the Lyman-α forest structures. These choices allow us to compare differences due to power spectrum shape and normalization and growth rates separately.

The model parameters are: an open cold dark matter (OCDM), $h = 0.7$, $\Omega_{nr} = 0.3689$ and $\sigma_8 = 0.81$: a flat non-zero cosmological constant model (ΛCDM), $h = 0.7$, $\Omega_{nr} = 0.3349$ and $\sigma_8 = 0.93$: a hot-cold model (HCDM), $h = 0.5$, $\Omega_{\nu} = 0.2$, $\Omega_{cdm} = 0.75$ and $\sigma_8 = 0.82$ (2 degenerate neutrino species) and 3 standard (SCDM) models with $h = 0.5$, $\Omega_{nr} = 1.0$ and $\sigma_8 = 0.44$, 0.67 and 1.0. $\Omega_{nr}$ refers to the critical mass fraction in non-relativistic matter (basically cold dark matter and baryons). $\Omega_{cdm}$ refers to the cold dark matter and $\Omega_{\nu}$ refers to hot dark matter (assumed to be neutrinos with 2 degenerate massive species) at $z = 0$. Figure 3.1 demonstrates the power spectra associated with these choices. The OCDM, ΛCDM, HCDM and
$\sigma_8 = 1.0$ SCDM models are consistent with microwave background observations. The OCDM, ACDM, HCDM and $\sigma_8 = 0.67$ SCDM models are broadly consistent with cluster constraints. At smaller scales, galaxy correlations (e.g. from Las Campanas: Lin et al. (1996)) are also consistent with these models. The $\sigma_8 = 0.44$ SCDM model is under-normalized with respect to all the constraints.

Fig. 3.1.— Power spectra for the four cosmologies. The top panel is at redshift $z = 0$. The effect of a top hat filter with scale $8\, h^{-1}\text{Mpc}$ is shown. The lower panel is scaled to redshift $z = 3$. The effect of a Gaussian filter with scale $0.5\, \text{Mpc}$ is shown. This was the basis for the normalization choices. The $\sigma_8 = 0.67$ SCDM model is shown.
3. Motivation for Constrained Calculations

To focus on complex regions, the standard periodic box approach to cosmological simulations is not appropriate. We do not want to simulate a large box to capture one object. We constrain our initial conditions to produce the objects we want in the middle of our box. Our simulations are not periodic and include complex modelling of the tides and structural environment at lower, but sufficient, resolution. Thus, we do not waste any computational effort. This results in two great advantages for focussed simulations: good mass resolution and good $k$-space sampling.

Good mass resolution allows us to concentrate on the scales needed to adequately treat the objects that form. For the Lyman-$\alpha$ forest, we wish to resolve dwarf galaxies and need an initial grid spacing $a_L = a_{\text{Lattice, High Res}} = 100$ kpc comoving. For our Lagrangian method this translates into high numerical resolution ($h_{\text{sph.min.}}$, $h_{\text{grav}} \sim 1$ kpc).

Good $k$-space sampling is also possible with non-periodic simulations. The competing demands of $k$-space sampling and resolution are further described in figure 3.2. Our method allows high resolution without compromising our long wave coverage by going beyond grid based FFTs, with a FastFT for high $k$ (which kicks out well before the fundamental mode is reached) that is superseded by two direct FTs; power-law then log $k$ sampling. Transitions between the methods occur so as to minimize the volume per mode in $k$-space. Well-sampled $k$-space is especially important for Ly$\alpha$ cloud and galaxy formation as opposed to cluster formation because the density power spectrum for viable hierarchical theories has nearly equal power per decade (approaching flatness in figure 3.2): if just the FFT is used, as is often the case in cosmology even for non-periodic calculations because there is only one fundamental $k$-mode along each box axis, the large scale structure in the simulation will be poorly modelled. This can also have a deleterious effect on small scale structure.

There is no point adding long waves without maintaining accurate large scale
tides and shearing fields during the calculation. We achieve this with a high resolution region of interest (grid spacing $a_L$, $50^3$ sphere) that sits within a medium resolution region ($2a_L$, $40^3$), in turn within a low resolution region ($4a_L$, $32^3$). The influence of ultra-long waves is included by measuring the mean external tide acting on the low resolution region in the initial conditions, adopting simple models for the ultra-long wave dynamics based on that measurement (e.g. linear, Zel'dovich, or homogeneous ellipsoid, as for peak patches (Bond & Myers 1996)) and applying the tide as an "external force" throughout the simulation. Linear ultra-long wave dynamics were adequate for this work because the largest scales did not go non-linear.

Small simulations must use non-periodic initial conditions or be forced into an artificial regime with low shear on the scale of the box. For a large enough periodic box the density perturbation on the scale of the box can be made small enough that the large scale shear is unimportant. A periodic box also has average density equal to the mean. The result is that for periodic initial conditions, large regions must be simulated to have the correct tides and density fall-off from overdense regions such as that shown in figure 3.6. Our technique of linear tides is more computationally efficient than large periodic boxes.

Simulating a large number of controlled patches in parallel is a form of adaptive refinement. In single simulations with adaptively refined regions, the initial conditions are not augmented with higher frequency waves, so the Lagrangian (i.e. mass) resolution remains fixed even though the Eulerian resolution may be superb. (This is especially vexing for voids.) When we refine a region by creating a high resolution realization with the information contained in peak patches, we optimally resample $k$-space to generate a new set of high frequency waves. It is clear that the cosmological codes of the future will have to simultaneously adapt in Eulerian and $k$ space, and the techniques explored here offer a promising path towards this goal.
Fig. 3.2.— The two (linear) power spectra shown (scaled to redshift 3) The upper curve at high \( k \) is the standard untilted CDM model, but normalized to cluster abundances, \( \sigma_8 = 0.67 \). The other has the same cosmological age (13 Gyr) and \( \Omega_B h^2 = 0.0125 \) but \( H_0 = 70 \) with \( \Omega_\Lambda = 0.67 \), slightly tilted to be COBE-normalized (\( n_s = 0.94 \)). (Also shown is a COBE-normalized CDM model, which misses the solid data point in the cluster-band (constraint from \( d\eta/d\chi \)).) The bands in comoving wavenumber probed by various simulations are contrasted. Periodic simulations may use the entire volume, but the \( k \)-space restriction to lie between the fundamental mode (low-\( k \) boundary line) and the Nyquist wavenumber (high-\( k \) boundary line) in the IC can severely curtail the rare events in the medium that observations especially probe, and prevents tidal distortions of the simulation volume. We use 3 \( k \)-space sampling procedures (FFT and two direct FTs) with the boundaries defined by which has the smallest volume per \( k \)-mode. Even though a 256\(^3\) Fourier transform was used, notice how early the direct sampling takes over (with only 10000 modes). Using an FFT with the very flat spectra in the dwarf galaxy (dG) band can give misleading results. The 3 low-\( k \) lines shown for our Ly\( \alpha \) and galaxy simulations correspond to the high, medium and low resolution fundamental modes. We actually include modes in the entire hatched region, with the tidal fields associated with the longer waves included by a self-consistent uniform tide on the LR simulation volume. \( h_m \) denotes our best resolution. \( a_\perp \) denotes the physical (best) lattice spacing for the grid-based Eulerian hydro codes of Cen and Zhang et al. \( (z = 3) \). \( k \)-space domains for two large scale structure \( (z = 0) \) simulations are also shown, a Klypin 256\(^3\) \( PM \) calculation and a Couchman (labelled \textit{hugh}) 128\(^3\) \( P^3M \) simulation. (see also Bond & Wadsley (1996)).
4. Structure Formation

In this section, we mathematically describe the formation of objects such as galaxies from initial fluctuations in the density field. The basic concept is of convergent mass flows smoothed on the mass scale of the final object. The rate of flow will not, in general, be isotropic in the proto-object patch of space. This introduces the use of strain or shear to characterize the anisotropies in the flow.

Generally, for a cold medium, there is a full non-linear map: \( \mathbf{x}(\mathbf{r}, t) \equiv \mathbf{r} - \mathbf{s}(\mathbf{r}, t) \), from Lagrangian (initial state) space, \( \mathbf{r} \), to Eulerian (final state) space, \( \mathbf{x} \). The map becomes multivalued as nonlinearity develops in the medium. It is conceptually useful to split the displacement field, \( \mathbf{s} = \mathbf{s}_b + \mathbf{s}_f \), into a smooth quasilinear long wavelength piece \( \mathbf{s}_b \) and a residual highly nonlinear fluctuating field \( \mathbf{s}_f \). If the \( \text{rms} \) density fluctuations smoothed on scale \( R_b \), \( \sigma(R_b) \), are \( < \mathcal{O}(1/2) \), the \( \mathbf{s}_b \)-map is one-to-one (single-stream) except at the rarest high density spots. If the medium is examined at sufficiently early times, fluctuations on all scales of interest are in the linear regime.

In the peak patch approach, \( R_b \) is adaptive, allowing dynamically hot regions like protoclusters to have large smoothing and cool regions like voids to have small smoothing. If \( D(t) \) is the linear growth factor, then \( \mathbf{s}_b = D(t)\mathbf{s}_b(\mathbf{r}, 0) \) describes Lagrangian linear perturbation theory, i.e. the Zeldovich approximation. The large scale peculiar velocity is \( V_{pb} = -a(t)\dot{s}_b(\mathbf{r}, t) \). \( a = 1/(1 + z) \) is the scale factor of the universe.

Derivatives of the background displacement field determine whether material is clumping and forming objects or diluting relative to the general expansion. Large values of displacement and thus peculiar velocities only imply a bulk flow. Important for us is the strain field (or deformation tensor):

\[
e_{b,ij}(\mathbf{r}) \equiv -\frac{1}{2} \left( \frac{\partial \dot{s}_bi}{\partial r_j} + \frac{\partial \dot{s}_bj}{\partial r_i} \right)(\mathbf{r}).
\]  

\[ (3.6) \]
It is useful to rotate the tensor to its principle axes, expressing it in terms of its eigenvalues,

\[ e_{b,ij}(\bm{r}) = - \sum_{A=1}^{3} \lambda_{\nu A} \hat{n}_{\nu A} \hat{n}_{\nu A}. \]  

(3.7)

\( \delta_{Lb} \equiv (\rho_b - \langle \rho \rangle) / \langle \rho \rangle = -e_{b,i}^* \) is the smoothed linear overdensity. The deformation eigenvalues are ordered according to \( \lambda_{\nu 3} \geq \lambda_{\nu 2} \geq \lambda_{\nu 1} \) and the \( \hat{n}_{\nu A} \) denote the unit vectors of the principal axes. In this system, \( x_{A} = r_{A}(1 - \lambda_{\nu A}(\bm{r}, t)) \) describes the local evolution. The Zel'dovich-mapped overdensity is,

\[ (1 + \delta_{Z})(\bm{r}, t) = |(1 - D(t)\lambda_{\nu 3})(1 - D(t)\lambda_{\nu 2})(1 - D(t)\lambda_{\nu 1})|^{-1}. \]  

(3.8)

A fold caustic forms when the largest eigenvalue \( D(t)\lambda_{\nu 3} \) reaches unity. In a Zeldovich map, a pancake develops along the surface \( \hat{n}_{\nu 3} \cdot \nabla r \lambda_{\nu 3} = 0 \).

The strain tensor is related to the peculiar linear tidal tensor by,

\[ \frac{\partial^2 \Phi_{P}}{\partial x^i \partial x^j} = -4\pi G \rho_{m} \bar{a}^2 e_{b,ij}. \]  

(3.9)

where \( \Phi_{P} \) is the peculiar gravitational potential. The linear shear tensor is simply \( \dot{e}_{b,ij} \). The anisotropic part of the shear or strain tensor has two independent parameters, the ellipticity \( e_{\nu} \) (always positive) and the prolaticity \( p_{\nu} \). The eigenvalues of the shear tensor can be expressed using these parameters as follows.

\[ \lambda_{\nu 3} = \frac{\delta}{3} (1 + 3e_{\nu} + p_{\nu}), \quad \lambda_{\nu 2} = \frac{\delta}{3} (1 - 2p_{\nu}), \quad \lambda_{\nu 1} = \frac{\delta}{3} (1 + 3e_{\nu} - p_{\nu}). \]  

(3.10)

5. **Patch statistics**

We return to the Gaussian random field description of the initial density field to discuss distributions for the related variables.

The density fluctuation level relative to the root mean square \( \sigma_{R_{\nu}} \), for the filter
scale $R_f$, defines the background or peak height (centred on the origin).

$$
\nu_{R_f} = \frac{1}{\sigma_{R_f}} \int \delta(r) W_G(r, R_f) dr.
$$

(3.11)

\( \nu \) is Gaussian-distributed with zero mean and a standard deviation of 1. Thus \( \nu \gtrsim 2 \) characterizes a notable peak and \( \nu \lesssim -2 \) a void on the scale \( R_f \).

Other properties of interest of the field include the gradient, \( \nabla_i \delta \) and the second derivative tensor \( \nabla_i \nabla_j \delta \). Bardeen, Bond, Kaiser & Szalay (1986) rotate the tensor to its principle axes and transform to a set of 10 independent variables. There exists a negative correlations between \( F \) and \( \nabla^2 \delta \), and thus among the 10 variables only 2 are correlated. The probability distribution for the set may be simply expressed.

The most important parameter for us is the shear. Thus, we must add the strain tensor \( e_{ij}(r) \), defined in equation 3.6, to the set of variables. The \( \nabla_i \nabla_j \delta \) are correlated with the \( e_{ij} \). The joint probability distribution for the now 15 independent variables can also be straightforwardly expressed.

The strain or deformation tensor describes the shape of an evolving region. Rotating it to its principle axes and finding the eigenvalues gives the ellipticity \( e_v \) and prolateness \( p_v \) of the ellipsoidal region, defined relative to the eigenvalues of the strain tensor in equation 3.10. The principle axes directions are given the three Euler angle variables which have no correlations, all orientations of the ellipsoid being equally probable.

Knowledge of the smoothed strain tensor, to first order, is all we need to predict the evolution of a patch. In the frame moving with the patch, the principle axes collapse in order (if the corresponding strain tensor eigenvalues are positive). The linear density evolution is given by the Zel'dovich map (equation 3.8). Bond & Myers (1996) employ an homogeneous ellipsoid model initialized with the shear parameters to predict the fate of peak patches.

We have \( e_v > 0 \) and \( |p_v| < e_v \) from the ordering \( \lambda_1 > \lambda_2 > \lambda_3 \). A peak would
require additional constraints to ensure all the $\lambda_i$ are positive. General patches do not have to satisfy such a constraint. We choose to deal in the variables $y_v = \nu e_v$ and $z_v = \nu p_v$ which remain usable when $\nu = 0$. The probability distribution for $y_v$ and $z_v$ is independent of $\nu$ and given by (Bond & Myers 1997).

$$P(y_v, z_v) = \frac{450}{8} \sqrt{\frac{10}{\pi}} 2 y_v (y_v^2 - z_v^2)e^{-15/2 y_v^2 - 5/2 z_v^2} dy_v dz_v. \quad (3.12)$$

$$P(y_v) = 225 \sqrt{\frac{10}{\pi}} (4 y_v^2 e^{-10 y_v^2}) + \sqrt{\frac{8}{125\pi}} (-y_v + 3 y_v^3) \text{Erf} \left( \sqrt{\frac{5}{2}} y_v \right) e^{-15/2 y_v^2} dy_v. \quad (3.13)$$

These distributions are shown in figure 3.3. The mean value for $y_v$, independent of $\nu$, is 0.535237. As can be seen from the figure, values such as $y_v > 1$ are extremely unlikely, as is near-zero shear $y_v \sim 1$. For peaks, $y_v$ is strongly dependent on $\nu$. In particular rarer peaks tend to be increasingly spherical with smaller values of $y_v$ and a non-zero mean for $z_v$. A peak strictly has a local maximum at the central point. High $\nu$ regions have a positive fluctuation that is not necessarily a peak.

Often, we consider constraints on multiple scale and/or desire to know what the environment of a single constraint is like. Given that a fluctuation of height $\nu_1$ exists at a point with a filter scale of $R_1$, it is useful to know the mean peak height $\nu_2$ at another filter scale $R_2$. From BBKS,

$$< \nu_2 | \nu_1 > = \epsilon \nu_1,$$

$$\epsilon = \frac{\sigma_{12}}{\sigma_{R_1} \sigma_{R_2}},$$

$$\sigma_{12}^2 = \int \overline{W_{R_1}(k)} \overline{W_{R_2}(k)} P(k) \frac{d^3 k}{(2\pi)^3}. \quad (3.14)$$

The mean shear on different scales has the same fixed amplitude in terms of $y_v$ and $z_v$, but the principle axes may change, though not very much on average: The alignment is highly correlated.
Fig. 3.3.— The probability distribution for the anisotropic shear parameters, $y_e$ and $z_e$. These distributions are different for regions constrained to be local extrema (e.g., a peak) but not if the mean background perturbation level is varied.
6. Designing Lyman-α Forest Simulations

There are strong constraints that determine the usefulness of simulations of the Lyman alpha forest. High resolution and sufficient size to properly capture the scales that are important at each redshift are simultaneously necessary. Fluctuations on a given scale are important if their amplitude is significant (say \( \geq 0.1 \)). If fluctuations on the scale of the entire simulation box collapse – the simulation is flawed. If scales that are a large fraction of a periodic box are collapsing then they will be modelled incorrectly because of missing tides and large scale power.

As discussed in Chapter One, the Jeans length at temperatures typical of a photo-ionized gaseous medium indicates that gas pressure will tend to inhibit the formation of baryonic structures with scales less than \( \sim 140 \text{kpc} \) comoving at \( z = 3 \). This prompted us to select 0.1 Mpc as our initial resolution. We have performed simulations indicating that poorer resolution does not adequately model the small structures responsible for the Lyman-α forest and that our choice has sufficient resolution for the problem (Chapter Four).

With \( 50^3 \) particles, our high resolution region is 5 Mpc across. For standard CDM with \( \sigma_8 = 0.667 \) (cluster normalization) an rms fluctuation (smoothed \( \delta = \delta \rho / \langle \rho \rangle \)) on a scale of 0.5 Mpc comoving (smoothed with a Gaussian filter \( \exp(- (r/R_f)^2) \)) will have a linear amplitude of \( \sigma_{G,0.5} = 1.05 \) at \( z = 3 \) and 1.4 at \( z = 2 \). By stopping our simulation at \( z = 2 \) we ensure typical fluctuations on this scale will just be collapsing. (For the other cosmologies studied, we retained \( \sigma_{G,0.5} = 1.05 \) at \( z = 3 \) as the normalization for the fluctuation spectrum. This choice is consistent with microwave background and cluster constraints for our open, vacuum and hot-cold dark matter models.)

We also perform simulations with higher normalizations \( \sigma_8 \). The degree of collapse is great at redshift \( z = 2 \). On a Gaussian filtered scale of 1.5 Mpc, the rms linear fluctuation amplitude for standard CDM is \( \sigma_{G,1.5} = 0.76(\sigma_8/0.67)(3/(1 + z)) \). At \( z = 2 \) with \( \sigma_8 = 1.0 \) the linear amplitude of a \( \nu = 1.1 \) fluctuation on a Gaussian
filter scale of $R_f = 1.5 \text{ Mpc}$ (corresponding to $\nu = 1.4$ for $R_f = 0.5 \text{ Mpc}$) at $z = 2$ is 1.25 – entering the final stages of collapse. This simulation is pushing the limits of our approach.

Rather than attempting to represent the universe with a single calculation, we perform a set. We use importance sampling of shearing patches (patches with the smoothed shear tensor prescribed at the centre) to maximize the statistical information we can get from a crafted set of relatively modest constrained-field SPH calculations, defined by a set of control parameters: the central $\nu R_f$, $e_v$, $p_v$ smoothed over a galactic-scale $R_f = 0.5 \text{ Mpc}$. We use a set of 5 values for the fluctuation level for the central constraint, $\nu = -1.4$, $-0.7$, $0$, $+0.7$ and $+1.4$ with $R_f = 0.5 \text{ Mpc}$. This allows us to sample rare peak and void patches, difficult to sample even in large box simulations (especially if FFTs are used). These are in addition to patches with more typical rms density contrasts. The values of $\nu$ and $e_v$ are sufficient to determine the properties of collapsed peaks. We can combine the results to get the frequency distribution of, say, $N_{HI}$ for a random patch of the Universe using Bayes theorem, which decomposes it into the frequency distribution for $N_{HI}$ for our constrained patches given the control parameters, measured from the simulations, and the known probability distribution of the control parameters: schematically.

$$P(\text{random - patch}) = \int P(\text{constr - patch|control}) \ P(\text{control}) \ d\text{control}.$$

We elected to use the mean value of the shear for each simulation: $y_v = \nu e_v = 0.535237$, $z_v = 0$. This is representative as long as the collapse is not severe on the scale of the constraint. This may be expected to fail for rarer peaks. We have investigated this by simulating an extremely sheared $\nu = 1.4$ peak ($y_v = 0.8$, $z_v = 0$). The modification from the mean $\nu = 1.4$, $y_v = 0.54$, $z_v = 0$ peak is substantial, but it does not grossly affect our results because $y_v = 0.8$ is substantially more rare and the weighting for all patches $\nu \gtrsim 1.4$ is already only 6.9%. The small cross-section in real space reduces the contribution at $z = 3$ to 2.6%. Of more concern is the effect of shear on our most likely case, $\nu = 0$. We demonstrate the impact of modifying the shear parameters on the column density distribution.
results in figure 3.4. We are confident that the mean shear is representative for the dominant cases.

For weak background fluctuations. the density structure nearby is important and should be constrained so that random variations in it do not dominate the result. To ensure that the region around our inner constraint would be typical, we applied additional constraints at $R_f = 1.0$ and $1.5 \, \text{Mpc}$ that the $\nu$ values take the mean value given the inner constraint (using equation 3.14). The mean shear value was used for these additional constraints, with the same alignment of the principle axes as the innermost constraint.

Aside from the constraints, all other parameters of the simulations were held constant - including the phases of the waves. This ensured that all differences were due to the background field. It also allowed detailed visual comparisons and examination of the formation of objects with different peak or void backgrounds - the biasing process.

For random patches $\nu$ is Gaussian distributed. thus a 2-$\sigma$ or higher background would occur only 0.22% of the time. There is an addition reason besides rarity that allows us to overlook high fluctuations. The patch probabilities derived from properties of Gaussian random fields apply to Lagrangian space. The final real (Eulerian) size of the patch is very much smaller for peaks and expanded for underdense patches. Our interest is focused on the Lyman-$\alpha$ forest - for which high peaks are not specifically necessary. Thus, the collapse of high peaks acts to further reduce their importance to the sample by reducing the cross-section for intersecting a QSO line of sight. Very rare peaks are effectively galaxies. They are important for damped Lyman-$\alpha$ absorption but not for the Lyman-$\alpha$ forest. The outskirts of galaxy peaks have larger than average $\nu$ values and are included in the overall statistics of $\nu$. Thus we model this contribution appropriately with our high $\nu$ patches.

The visual effect of $\nu$ variation is shown in figure 3.5. The same filamentary
Fig. 3.4.— The effect of shear on the distribution of column densities. The effect is greatest for overdense regions ($\nu = 1.4$) that make a small contribution to the final result. The mean density patch ($\nu = 0$) experiences minimal effects. The boxes shown are observed data and are intended as a guide for the shape of the curves. The normalization is close to correct. (The data should be compared to the weighted sum of all $\nu$'s – resulting in a curve slightly below the $\nu = 0$ result due to the greater contribution of voids. See Chapter Four)
structure and larger objects form in each case due to the presence of the same wave phases. The amplitudes are adjusted to satisfy the constraints but this has a very limited effect on the small scale structures. The background field acts to speed the evolution and collapse of small perturbations, enhancing both the filamentariness and the number density of collapsed halos in the \( \nu > 0 \) cases. The filaments are dramatically suppressed for \( \nu = -0.7 \). The typical collapsed objects visible have associated total masses of order \( 10^9 - 10^{11} \) solar masses and rotation speeds in the range 30 – 100 km s\(^{-1}\). All panels are for a standard initially scale-invariant \( \Omega_m=1 \) CDM model with \( \Omega_B=0.05 \), \( h=0.5 \), normalized to \( \sigma_8=0.67 \).

The simulation results are discussed in detail in Chapter Four, specifically with regard to the implications for the Lyman-\( \alpha \) forest. We also performed 5 galaxy scale simulations, parameterized by the same \( \nu \) set, for the SCDM case. These had a high resolution region 15.625 Mpc comoving across and an initial lattice spacing of 0.3125 Mpc comoving. They were constructed in exactly the same fashion as the Lyman-\( \alpha \) scale simulations and thus represent a statistically complete sample. We use them to demonstrate the effects of poor resolution.

### 7. Peak Patches

We briefly describe the Bond & Myers (1996) peak patch method. The technique identifies candidate peak points using a hierarchy of smoothing operations on the linear density field \( \delta_L \). To determine patch size and mass we use an ellipsoid model for the internal patch dynamics, which are very sensitive to the external tidal field. A byproduct of the ellipsoid model is the internal (binding) energy of the patch and the orientation of the principal axes of the tidal tensor. We apply an exclusion algorithm to prevent peak-patch overlap. For the external dynamics of the patch, we use a Zel’dovich map with a locally adaptive filter \( R_{pk} \) to find the velocity \( V_{pk} \) (with quadratic corrections sometimes needed). The peaks are rank-ordered by mass (or internal energy). Thus, for any given region, we have a list of the most important
Fig. 3.5.— These particle plots show the result of changing background field height $\nu$. Clockwise from the top left the $\nu$ values are 0.0, -0.7, 0.7 and 1.4 respectively. The underlying waves are the same and thus similar structures are present in each case. Note the huge enhancement of structure and dwarf objects in the peak cases $\nu > 0$. The panels show the inner 2.5 Mpc of the simulations at $z = 3$. The shades indicate increased $n_{HI}$ densities, ranging from $10^{-12}$ (grey) to $10^{-5}$ (white) cm$^{-2}$. The peak values in the simulations are around $\sim 1$ cm$^{-2}$.
peaks. Using the negative of the density field, we can also find void-patches. First among the virtues of the method is that it represents a natural generalization of the Press-Schechter method to include non-local effects. The hugely popular and trivial to implement, Press-Schechter (1976) method for determining \( n(M) \) has been the principal competitor to the peaks theory over the years, but it has no real physical basis and disagrees strongly with the spatial distribution (Bond et al. 1991). Its predictions for \( n(M) \) fit those of \( N \)-body group catalogues well, however. Peak patches is a natural generalization of BBKS single-filter peaks theory to allow a mass spectrum and solve the cloud-in-cloud (i.e., peak-in-peak) problem. It allows efficient Monte Carlo constructions of 3D catalogues and gives very good agreement with \( N \)-body groups. It uses an accurate analytic theory to estimate peak properties, \((e.g., \text{ mass and binding energy from mean-profiles, using } d_{L, \text{crit}}(e_v), \langle e_v | \nu_{pk} \rangle)\). \( \nu_{pk} \) is the height of the density peak with respect to the rms for that scale. Peak patch handles merging, with high redshift peaks being absorbed into low redshift ones.

8. **Crafting High Resolution Simulations**

Although it is usual to evolve ambient "random" patches of the Universe in cosmology, there are obvious advantages to spending one's computational effort on the regions of most interest. Single peak constraints are very useful if cluster or galaxy formation is the focus, while multiple peak constraints are more useful if superclusters, cluster substructure, or filaments and walls are the focus. The essential features at a given epoch of the filamentary structure and the wall-like webbing between the filaments are largely defined by the dominant collapsed structures and the peak patches that gave rise to them. A general method for building peak environments is suggested: construct random field initial conditions that require the field to have prescribed values of the peak shear (smoothed over the peak size at the peak position), for a subset of the size-ordered list of peaks that will have a strong impact on the patch to be simulated. Only a handful \( N_{pk} \) of peaks and/or voids is usually needed to determine the large-scale features, effectively compressing the
To illustrate how this works, we created a random (unconstrained) initial state for a CDM model in a 40 Mpc box, our “galaxy” simulation. We found peak patches according to the method described in Bond & Myers (1996) and focused on a specific subregion exhibiting a strong filament, choosing the peaks and voids that were expected to exert strong tidal influences within and upon the patch, as described in figure 3.6. The region chosen was just above the large central cluster of peaks in the top left panel of figure 3.6. We constructed a higher resolution set of “Lyα cloud” initial conditions for this patch, which the “galaxy” initial condition simulation could not resolve well enough to address the low column depth Lyα forest of interest to us.

By compressing the initial data in our target region to just the positions and shears of a few large peak patches and then forming a constrained realization and applying different random waves (optimally-sampled for the smaller region) than the original 40 Mpc initial condition used, we know we will get high frequency structure wrong. But clearly the large scale features are the same. This is in spite of the tremendously complex filamentary structure just below our chosen sub-region. We chose the peaks on the basis of rareness (size) and proximity to the patch (using an algorithm roughly based on correlation function falloff from each peak). The five peak-patches used for this companion Lyα simulation had the following masses (in units of $10^{11} M_\odot$) and halo velocity dispersion (in units of km s$^{-1}$) as determined from the binding energy: 3.6, 77; 3.5, 80; 1.4, 57; 0.85, 48; 0.51, 40. These agree well with what our group finder finds in the simulation at this redshift. The two void-patches used had Lagrangian masses of 2.4 and 0.72, and were outside the high resolution interior. The approximate alignments of the shear tensors for the peak patches inside ensure that a strong filament exists. figure 3.7 shows how the filament looks in HI column density.

9. The Cosmic Web
In 1965, Lin, Mestel & Shu showed that a cold triaxial collapse implied an oblate "pancake" would form. In 1970, Zel'dovich developed his famous approximation and argued that pancakes would be the first structures to form in the adiabatic baryon-dominated universes popular at that time.

Doroshkevich (1973) and later Doroshkevich & Shandarin (1978) were among the first to apply the statistics of Gaussian random fields to cosmology, in particular $\lambda_{vA}$, at random points in the medium. Arnold, Shandarin and Zeldovich (1982) made the important step of applying the catastrophe theory of caustics to structure formation. Their work suggested the following formation sequence: pancakes first, followed by filaments and then clusters. This sequence should be compared to the Bond, Kofman & Pogosyan (1996) Cosmic Web picture formation sequence: clusters first, followed by filaments and then walls. They also showed that filaments bridging cluster peaks are really ribbons, walls are webbing between filaments in cluster complexes, and that walls are not really classical pancakes. For the Universe at $z \sim 3$, massive galaxies play the role of clusters, and for the Universe at $z \sim 5$ more modest dwarf galaxies take on that role.

The Web story relies heavily upon the theory of Gaussian random fields as applied to the rare "events" in the medium, e.g., high density peaks. Salient steps in the development of this theory begin with Bardeen et al. (1986, BBKS), where the statistics of peaks were applied to clusters and galaxies, e.g. the calculation of the peak-peak correlation function, $\xi_{pk,pk}$. In a series of papers, Bond (1986-90) and Bond & Myers (1990-93) (see references in Bond & Myers (1996)) developed the theory so that it could calculate the mass function, $n(M)dM$. The theory was also applied to the study of how shear affects cluster alignments (e.g. the Binggeli effect whereby nearby clusters tend to have their axes aligned), and to Ly$\alpha$ clouds, 'Great Attractors', giant 'cluster-patches', galaxy, group and cluster distributions, dusty PGs, CMB maps and quasars. This culminated in the Bond & Myers "Peak-Patch Picture of Cosmic Catalogues"; i.e. peak patches as a descriptive and predictive theoretical tool for the statistics of galaxy populations.
Bond, Kofman & Pogosyan (1996) concentrated on the impact the peak-patches would have on their environment and how this can be used to understand structure formation. They showed that the final-state filament-dominated web is present in the initial conditions in the $\delta_{L,6}$ pattern, a pattern largely determined by the position and primordial tidal fields of rare events. They also showed how 2-point rare-event constraints define filament sizes. The strongest filaments are between close peaks whose tidal tensors are nearly aligned. Strong filaments extend only over a few Lagrangian radii of the peaks they connect. They are so visually impressive in Eulerian space because the peaks have collapsed by about 5 in radius, leaving the long bridge between them, whose transverse dimensions have also decreased. This is illustrated by the lower right panel of figure 3.6 in which the aligned galaxy peaks are connected by strong filaments. Strong vertical filaments are a product of the vertical alignment of the peaks' tidal tensors, which simultaneously acts to prevent a strong horizontal filament between the top two peaks. The reason for this phenomenon is that the high degree of constructive interference of the density waves required to make the rare peak-patches, and to preferentially orient them along one axis, leads to a slower decoherence along that axis than along the others, and thus a higher density. 3-point and higher rare-event constraints of nearby peaks determine the nature of the webbing between the filaments, also evident in figure 3.6.

The Cosmic Web picture predicts the basic structural components of the IGM as a function of scale, epoch and cosmology. For density contrasts $\delta \geq 100$, the rare events at $z \sim 3$ are massive galaxies, observed as damped Ly$\alpha$ absorbers, while the typical collapsed objects are dwarf galaxies responsible for Lyman Limit and metal line systems. The medium is visually dominated by $\delta \sim 5 - 10$ filaments, bridging massive galaxy peaks, with smaller scale filaments within the larger scale ones bridging smaller dwarf galaxies contributing the most to the $N_{HI} \lesssim 10^{14.5}$ Ly$\alpha$ forest.
These plots show the reconstruction of a galaxy-galaxy filament present in our 40 Mpc "galaxy" simulation. The cosmology is a standard, initially scale-invariant $\Omega_{m}=1$ CDM model with $\Omega_B=0.05$, $h=0.5$, normalized to $\sigma_8=0.67$. We used the cosmological SPH+P^3MultiGrid code described in Chapter Two. We identified the peak-patches that should collapse by $z = 4$ in the initial conditions (IC), as shown in the top left panel (which is $10 h^{-1}\text{Mpc}$ across, comoving). For each patch, the outer ellipsoid represents the alignment of the shear tensor and the inner sphere is an estimate of the final object size. In the lower left panel, we zoom in on the central filamentary web structure and overlay dark matter from a simulation of these IC (panel 2.5 $h^{-1}\text{Mpc}$ across). The five peak patches (at $z = 4$ with overdensity 180) and two voids that define this region were used as constraints for a new higher resolution IC (12.8 Mpc), which we also evolved numerically. These peaks are shown overlaid with dark matter from the new simulation in the right hand panels. The top right panel is a different orientation to the others that shows the filament more clearly, also shown in $n_{HI}$ in figure 3.7. These panels demonstrate that peaks represent an excellent way to compress the essential information about large scale filamentary behaviour.
Fig. 3.7.— Left: A pseudo ray-traced rendering of the HI density for a region at $z=4$ constrained by the complex pattern of galaxy-scale peak and void shear-fields described in the text, encoding the main large scale structure features, and which ensures strong filamentary webbing arises in the patch. The colour scale codes the temperature variation, with the darker shades in the centre of objects representing high column cold gas. The region is $2.5\, h^{-1}\, Mpc$ across, comoving, and corresponds to the top right panel of figure 3.6. Right: $n_{HI}$ for an ambient patch of the Universe at $z=3$ with control parameter $v_b=0$, simulation as described in section 6. Note the filament dominance in both cases.
10. Peak Patches at High Redshift

$N$-body calculations cannot currently cover large enough volumes of space with sufficient resolution to simulate high redshift galaxy catalogues (with clustering) to compare with e.g. Steidel et al. (1997) cluster-like structure at $z \sim 3.1$. Only intermediate and small scale structures are accessible to direct gasdynamical simulations of the Lyα forest. A peak-patch catalogue with ultra-long waves included can treat such large volumes quickly and accurately. Galaxy halos are identified with regions of space determined to have collapsed using ellipsoidal internal dynamics. with external tidal fields playing a large role. Bulk properties like mass, binding energy, velocity and shear for the patches are subsequently measured and can be used with single-patch hydrodynamics or phenomenology to predict internal gas profiles.

Our simulation results indicate that objects dominate the contribution to Lyman-α absorption columns greater than $\sim 10^{16}$ cm$^{-2}$. With the knowledge afforded us from the Cosmic Web regarding the filamentary bridging structures it may be possible to model the associated lower column density systems. Focused constrained field initial conditions and simulations could build up a look-up table of filaments corresponding to peak pairs, or typical filaments could be constructed from mean profiles with a phenomenological approach.

In figure 3.8, we show the (impressively grand) large scale clustering in galaxies at various velocity dispersion cuts, in comoving space and in redshift space. Motivated by Steidel et al. (1997), we tiled regions from $z=2.8$ to 3.5 encompassing $18' \times 18'$ (in angle on the sky) with 128$^3$, 40 Mpc boxes. The size was chosen to resolve “dwarflet” peak-patches with binding energy $v_{BE} \sim 30$ km s$^{-1}$. Depending

<table>
<thead>
<tr>
<th>Cosmology</th>
<th>$\sigma_8$</th>
<th>$H_0$, Age</th>
<th>$\Omega_m$</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCDM</td>
<td>0.67</td>
<td>50, 13 Gyr</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>ΛCDM</td>
<td>0.91</td>
<td>70, 13 Gyr</td>
<td>0.335</td>
<td>$\Omega_\Lambda=0.665$, $n_s=1$</td>
</tr>
<tr>
<td>OCDM</td>
<td>0.91</td>
<td>70, 11 Gyr</td>
<td>0.37</td>
<td>$n_s=1$</td>
</tr>
<tr>
<td>HCDM</td>
<td>0.81</td>
<td>50, 13 Gyr</td>
<td>0.8</td>
<td>$\Omega_m=0.2$, 2 degenerate species</td>
</tr>
</tbody>
</table>
upon cosmology (Table 1), 15 to 22 boxes were needed. The cosmological parameters are slightly modified from those used for the Lyman-α forest simulations. Tiling for open (hyperbolic) universes is an interesting exercise. Optimal wavenumber sampling was used (as in section 8.), with phase-coherent ultra-long and short waves consistently joining box to box. The filaments bridging these dwarf galaxies define the lower $N_{HI}$ Lyα forest, while halos with $v_{BE} \sim 30 \text{ km s}^{-1}$ dominate the $N_{HI} \sim 10^{15}$ regime. The forest hydrodynamical simulations cover only 1/8 of the 40 Mpc at high resolution. Figure 3.8 therefore emphasizes that large scale waves must be included in our small scale “shear-patch” simulations, whose size is governed by our need to resolve 1 kpc structure in galaxies at $z=3$. The patches then typically include many $30 \text{ km s}^{-1}$ halos but only a handful of $90 \text{ km s}^{-1}$ ones. The currently largest periodic simulations with this resolution are not much bigger, but miss the long waves. We can use these peak-patch/cosmic web simulations to compare predictions for different cosmologies with the large scale structure probed by multiple quasar line-of-sight data and long range velocity space correlations in quasar spectra as well as emerging high-z catalogues.
Fig. 3.8.— The redshift-space (left) and comoving-space (right) distributions of galactic peak-patches. Top left is $18'$ wide and $9'$ thick, with a $v_{BE} > 290 \text{ km s}^{-1}$ cut, roughly corresponding to Steidel's damped Lyα systems Steidel et al. (1997). The 4 cosmologies shown, \{$S, \Lambda, O, H$\}CDM, all reveal large scale features, though more pronounced in OCDM and ACDM. Top right shows the spatial distribution, this time for bright galaxies with a $200 \text{ km s}^{-1}$ and $3 \times 10^{10} \text{ M}_{\odot}$ baryonic mass cut, for a region now $4.5'$ thick. Bottom left and right show a $90 \text{ km s}^{-1}$ cut, for a $1'$ thick slice.
REFERENCES


Steinmetz. M. 1996, to be published in The Early Universe with the VLT. ESO Workshop


Chapter 4

Results

1. The Inter-Galactic Medium

Here we describe the common elements to all the cosmological cases: A general picture of the IGM. In figures 4.1, 4.2 and 4.3 we show visualisations of the gaseous medium at high redshift. The dominant structures present in the medium are clearly filamentary. Another important structural element is dwarf galaxies. We investigate the relative importance of these elements and the degree to which we may ascribe features of Lyman-α absorption spectra to each in section 12. of this chapter. Other features present are walls between the filaments. These are quite tenuous but may be of importance for their Helium absorption.

The dark matter in all cases clumps without regard to pressure and, with sufficient resolution, is always more clumped than the gas. There are not always gas clumps corresponding to dark matter clumps. This is due to the gas pressure inhibiting structure formation as discussed earlier. Evolution of the medium can be described in terms of the lessening in importance of the filaments and other tenuous structures with the cosmic expansion and drainage into larger objects. This is shown visually when the four panels in each figure are compared. The difference between gas and dark matter evolution is apparent on comparison of figures 4.1 and 4.2. The
substructure present in the dark matter is effectively smoothed away in the gas. An interesting feature at lower redshifts is the fragmentation of the filaments in the dark matter to a much greater extent than in the gas. We expect the dominance of filamentary structures to subside as the typical fluctuation level rises and non-linear growth tends to focus the matter into objects. Material is also being arranged into larger scale filaments. The pressure of the gas allows it to retain its small scale filamentary character longer than the dark matter. The filamentary character of the universe as a function of scale is determined by the effective slope of the power spectrum at that scale. At Lyman-α scales the effective slope $n_{\text{eff}} \lesssim -2$, is highly conducive to dominant filamentary structures. As time progresses, larger scales structures assume importance and we move into a regime where filaments are less dominant (Bond et al. (1996)).

At very low redshifts we expect only objects to continue making important contributions to absorption. This is consistent with the high rate of evolution in the line counts seen at high redshift versus the slow evolution at low redshift. The fast evolution is because filamentary structures are tied to the cosmic expansion and thus evolve quickly to lower density, whereas virialized objects evolve independently. This result is strongly apparent in figure 4.3: The filamentary structures are rapidly evolving down in column density ($N_{\text{HI}} \lesssim 10^{15}$ at $z = 3$), but the gas near objects is retaining a fixed column. Objects evolve through merging. Because of the increasing importance of larger scales in this evolution we can not follow it below redshift $z = 2$ with our simulations. We would have to simulate larger volumes, both to achieve a good sample and to include the longer waves.

To construct our samples of the universe for each cosmology, we sampled the parameter space defined by the background fluctuation level in five simulations and given in terms of $\nu$. A constraint was applied to enforce the Gaussian smoothed fluctuation level with a filter scale of 0.5 Mpc comoving to be $\nu\sigma_{0.5}$, where $\sigma_{0.5}$ is the mean fluctuation on this scale, fixed at 1.05 for all the cosmologies. A single cosmology set of simulations is quite representative of the effect of changing $\nu$. 
Fig. 4.1.— Column density in baryonic matter through 2 Mpc thick slices for 4 redshifts. The cosmology is standard CDM with $\nu = 0$. The contours in $\log_{10} N_{\text{baryon}}$ compared to the mean column through a slab of this thickness are -0.75 (dark), -0.5 (dark grey), -0.25 (medium grey), 0 (light grey), 0.5 (dark), 1 (black) and 1.5 (white). One way to interpret these values is to consider that for an object with a physical scale of 25 kpc at $z = 3$, the $10^1$ (black) contour corresponds to a overdensity of 200 for that object. Thus the black contour is an indicator of collapsed, virialized objects.
Fig. 4.2.— Column density in dark matter, through 2 Mpc thick slices for 4 redshifts. The cosmology is standard CDM with $\nu = 0$. The contours in $\log_{10} N_{\text{dark matter}}$ compared to the mean column through a slab of this thickness are -0.75 (dark), -0.5 (dark grey), -0.25 (medium grey), 0 (light grey), 0.5 (dark), 1 (black) and 1.5 (white). One way to interpret these values is to consider that for an object with a physical scale of 25 kpc at $z = 3$, the $10^1$ (black) contour corresponds to a overdensity of 200 for that object. Thus the black contour is an indicator of collapsed, virialized objects.
Fig. 4.3.— Column density, $N_{HI}$, through 2 Mpc thick slices for 4 redshifts. The cosmology is standard CDM with $\nu = 0$. Note how the virialized objects maintain their density structure while the filamentary connections lower in density with the decreasing redshift. The contours in $\log_{10}N_{HI}$ are $11-12.75$ (dark), $13$ (dark grey), $13.5$ (medium grey), $14$ (light grey), $14.5$ (dark), $15$ (black) and $17$ (white).
Figure 4.4 visually demonstrates the results. The small scale waves are kept similar by virtue of identical wave phase choices. Note the huge enhancement in structure present in the peak case, \( \nu = 1.4 \). The number of dwarf galaxies is enormously increased over the mean case. By contrast, the void case, \( \nu = -1.4 \) has suppressed all but the highest peaks. Note also the huge expansion of the void case and the compression in the peak case. This effectively gives the void regions a large cross-section for intersecting a line-of-sight for QSO absorption. This weighting acts to reduce the number of absorption features in the spectra over the mean case, as shown later. Thus, the mean density case is not representative, showing the importance of non-periodic simulations.

We make a visual comparison between the cosmologies later (in section 4.), when the appropriate flux levels to fit the data have been determined for each cosmology.

Our set of simulations with different mean background fluctuations and the same high frequency wave structure provides an opportunity to demonstrate the process of biasing. Taking the \( \nu = 0 \) case (no mean background) and comparing it to the \( \nu = 0.7 \) case (a smooth background peak), the additional amplitude provided by the existence of the peak raises the small scale positive fluctuations at redshift \( z = 5 \) to a similar level to that they would achieve via linear growth at \( z = 3 \) in the absence of such a peak. The amplitude extrapolated to \( z = 0 \) provided by the peak is \( \delta \sim 4 \). Thus fluctuations hitting a linear growth amplitude of \( \delta = 1.6 \) at \( z = 3 \) will achieve this amplitude at \( z = 5.5 \) with the influence of the peak (and slightly later further off the peak). We compare \( z = 3 \) and \( z = 5 \) in figure 4.5, noting the similarity in the structures present. This similarity is much reduced in a direct comparison at the same redshift, \( z = 3 \), in figure 4.4.

1.1. Gas Properties
Fig. 4.4.— $N_{HI}$ Column density through 2 Mpc thick slices at $z = 3$. We show the effect of adjusting the background fluctuation level. Five simulations with $\nu = -1.4, -0.7, 0, +0.7$ and 1.4 comprise our sample for each cosmology. The cosmology shown here is standard CDM. The contours in $\log_{10} N_{HI}$ are 11-12.75 (dark), 13 (dark grey), 13.5 (medium grey), 14 (light grey), 14.5 (dark), 15 (black) and 17 (white).
Fig. 4.5.— 2 Mpc thick slices for a mean density patch at \( z = 3 \) (left) and a patch with enhanced background density but the same wave phases at \( z=5 \) (right). The cosmology is standard CDM. The top panels show log contours in dark matter density and the lower panels contours in column density, \( N_{HI} \). The contours in \( \log_{10} N_{HI} \) are 11-12.75 (dark), 13 (dark grey), 13.5 (medium grey), 14 (light grey), 14.5 (dark), 15 (black) and 17 (white). The similarity in the structure present is to be expected because similar linear overdensities indicate that similar objects should be collapsing in each case. The higher redshift case, however, has a lower Jeans mass. This is a visual demonstration of biasing – the existence of a background peak enhances the formation of smaller scale objects.
In figure 4.6 we show the general properties of the medium. The gaseous medium is well described by several phases. Underdense gas is expansion cooling out of equilibrium. Roughly mean density gas is accreting onto filaments and draining towards objects. Near objects the gas is often shocked above equilibrium to high temperatures. Gas tightly bound within objects can achieve sufficient density to rapidly cool and achieve equilibrium temperatures typical of the neutral ISM (10,000K). The UV heating creates an environment where the equilibrium temperature \( T_{\text{eqm}} \propto \rho^a \), with \( a \sim 0.7 \) at low densities and \( a \sim -0.2 \) at high densities. The gaseous medium is not susceptible to the formation of multiple phases via purely thermal instabilities. This statement can be made on the basis of the Field's criterion (1965), mathematically stated as \( \partial (\text{Heat} - \text{Cool})/\partial T \) at constant pressure is less than zero. Graphically, on the figure, the criterion would require that the equilibrium temperature curve have a slope \( a < -1 \). The cooling processes included in this work are described in detail in Chapter 2. We are working with an assumption of uniform zero metallicity – the addition of metals and the associated cooling would increase the likelihood of phase separation via instability. We know such processes occur within galactic halos. The degree of pollution of the medium by metals is an unresolved question not addressed here, but vital to interpret metal line observations (e.g. Hellsten et al. (1997)). These observations will allow us to constrain models of metal production and the evolution of dwarf galaxies. As discussed in Chapter 2 metallicity affects the cooling rates in the gas.

2. Spectra

To analyse our simulations we produce artificial spectra, with a signal to noise ratio of 100 and 5 km s\(^{-1}\) pixels. We experimented with adjusting the ionizing flux levels to values different to those used in the simulations. We always set the ion abundances to equilibrium values for the flux chosen and then create the spectra. We evaluate this rescaling in section 9.
Fig. 4.6.— The physical state of the gas. Results shown are for a standard cold dark matter simulation ($H_0 = 50\text{km}\text{s}^{-1}$, $\sigma_8 = 0.67$, $\Omega_0 = 0.05$). The upper panels are for $J_{-21} = 0.5$ and the lower for $J_{-21} = 0.2$. The other cosmologies look very similar. The curves give the equilibrium temperature at a given density, including the effects of expansion cooling for the $J_{-21} = 0.5$ case. The low density gas is out of equilibrium because of long cooling times and the intermediate density gas is shock heated above equilibrium (especially at $z=2$). The points shown are median values with the vertical bars marking 1st and 3rd quartiles for the bin.
We sample a regular spatial grid of lines of sight running along the three axial directions through our simulations. A typical line of sight is $\sim 500 \text{ km s}^{-1}$ long and a single simulation sample $\sim 10^6 \text{ km s}^{-1}$ at $z=3$. We add 1 percent noise which is comparable to HIRES Keck data. We have not attempted to match the detailed statistics of the observational noise.

A sample length of spectrum from our simulations is shown in figure 4.7, with Voigt profile fits also shown. Voigt fitting is described later. In figure 4.8 we show details of the gas that is responsible for creating the lines that are observed. In figure 4.9, we compare the same line of sight passing through each of the four cosmologies simulated. The same wave phases are present in each case, so the structures are very similar. The comparison can not be perfect as transverse motions will shift the lines of sight slightly for different cosmologies.

3. Weighting the Simulations

We need to define the useful region of each simulation for the purposes of measuring statistics. Our basic criterion is that absorption lines be produced in the inner high resolution region (5 Mpc for standard runs). Unfortunately, though the regions are distinct in physical space, in velocity space the border is fuzzy. Outer regions can have peculiar velocities that make lines due to outer material appear within the velocity space boundaries of the high resolution region. We define the edge in velocity space where the particles defining the high resolution boundary begin to dominate the contribution the lines. The boundary is Lagrangian, following the mass. The mass clusters, ending up collapsing into the objects. The boundary is thus more likely to end up in a collapsed region compared to a random line in physical (Eulerian) space through the simulation. There is thus a natural excess of lines near the boundary. Correctly including the contribution of the entire Lagrangian region means that we do not have the freedom to move arbitrarily inwards from the boundary. There is thus some contamination of the results by
lower resolution material adjacent to the boundary. This effect is not large and affects 10 – 20% of the volume.

Table 1: Weighting for High Resolution SCDM simulation.

<table>
<thead>
<tr>
<th>$\nu$</th>
<th>Erf</th>
<th>$Z=2$ $\Delta \nu$ W</th>
<th>$Z=3$ $\Delta \nu$ W</th>
<th>$\frac{dn}{dN_{HI}}$ $\Delta \nu$ W</th>
<th>$Z=4$ $\Delta \nu$ W</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1.4</td>
<td>0.0688</td>
<td>431 0.154</td>
<td>520 0.137 0.44</td>
<td>613 0.125</td>
<td></td>
</tr>
<tr>
<td>-0.7</td>
<td>0.2415</td>
<td>342 0.354</td>
<td>432 0.339 0.73</td>
<td>522 0.326</td>
<td></td>
</tr>
<tr>
<td>0.0</td>
<td>0.3794</td>
<td>254 0.341</td>
<td>343 0.358 1.21</td>
<td>432 0.367</td>
<td></td>
</tr>
<tr>
<td>0.7</td>
<td>0.2415</td>
<td>201 0.124</td>
<td>275 0.140 1.94</td>
<td>349 0.125</td>
<td></td>
</tr>
<tr>
<td>1.4</td>
<td>0.0688</td>
<td>192 0.026</td>
<td>224 0.025 2.72</td>
<td>274 0.028</td>
<td></td>
</tr>
</tbody>
</table>

Combining our 5 simulations to give an overall sample provides important insight into the contribution of different Lagrangian patches of space to any observational measurements. For many observations, only the rare peaks will make a contribution because they produce observable objects such as galaxies. For the Lyman-α forest, all space produces detectable absorption lines. The voids dominate the forest line sample, because the initial region expands in both physical and in velocity space. Its cross-section for being pierced by the line-of-sight to a quasar becomes much larger and a greater length of the spectrum is occupied per void-like region so pierced. The impact on our results is shown in table 1. Erf denotes the statistical Lagrangian weighting attributed to each patch. As $\nu$ (the simulation mean density measure) is Gaussian distributed, we estimate this weight by $\text{Erf}(\nu + \Delta \nu/2) - \text{Erf}(\nu - \Delta \nu/2)$. If anything, this scheme favours outlying $\nu$'s. $\Delta \nu$ denotes a mean velocity space width for the high resolution section of the simulations. The width through the centre of the box is double this. W gives the total weighting for the box. Peak regions are thus of limited importance for the forest. A typical contributing spatial patch might be described as having $\nu = -0.7 - 0.0$, and thus slightly void-like.

An interesting rough application of this data is to study the intrinsic variation between lines of sight due to cosmic variance. We have listed in the table $\frac{dn}{dN_{HI}}$, which is the differential count of $10^{13.85}\text{cm}^{-2}$ column lines. This bin is representative of the counts of low column lines around $\sim 10^{14}\text{cm}^{-2}$. We can estimate the variation
Table 2: Weighting for SCDM simulations at different resolutions. Z=3. HR and LR denote high (0.1 Mpc) and low (0.3 Mpc) resolution simulations. ($J_{-21} = 0.5$)

<table>
<thead>
<tr>
<th>$[h] \nu$</th>
<th>Erf</th>
<th>HR</th>
<th></th>
<th></th>
<th>LR</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$\Delta v$</td>
<td>$W$</td>
<td>$D_A$</td>
<td>$\Delta v$</td>
<td>$W$</td>
<td>$D_A$</td>
</tr>
<tr>
<td>-1.4</td>
<td>0.0688</td>
<td>520</td>
<td>0.137</td>
<td>0.110</td>
<td>1609</td>
<td>0.095</td>
<td>0.114</td>
</tr>
<tr>
<td>-0.7</td>
<td>0.2415</td>
<td>432</td>
<td>0.339</td>
<td>0.160</td>
<td>1476</td>
<td>0.286</td>
<td>0.131</td>
</tr>
<tr>
<td>0.0</td>
<td>0.3794</td>
<td>343</td>
<td>0.358</td>
<td>0.254</td>
<td>1305</td>
<td>0.375</td>
<td>0.153</td>
</tr>
<tr>
<td>0.7</td>
<td>0.2415</td>
<td>275</td>
<td>0.140</td>
<td>0.358</td>
<td>1169</td>
<td>0.197</td>
<td>0.174</td>
</tr>
<tr>
<td>1.4</td>
<td>0.0688</td>
<td>224</td>
<td>0.025</td>
<td>0.484</td>
<td>1016</td>
<td>0.045</td>
<td>0.202</td>
</tr>
<tr>
<td>Mean</td>
<td>368</td>
<td>0.219</td>
<td></td>
<td></td>
<td>1340</td>
<td></td>
<td>0.149</td>
</tr>
</tbody>
</table>

between two lines of sight with a simple Monte-Carlo simulation of 1,000 lines of sight. An observed spectrum uses the data between Lyman-α (1216 Å) and Lyman-β (1026 Å). This corresponds to a redshift interval of $\Delta z \sim 0.6$ at $z \sim 3$. For an SCDM universe this is 750 Mpc comoving in length, converting to 75,000 km s$^{-1}$ of Hubble flow, with a mean box width of 368 km s$^{-1}$. This translates to 203 high resolution boxes with $\nu$ values chosen according to the probabilities given in the table. This implies an rms variation of 3% in the line counts between spectra. This is a lower limit because adjacent boxes are not random but correlated. As we move to higher redshifts, the number of lines is much larger, so the variance is less. There is a corresponding 3% rms variation in the mean flux depression. Variance of this nature is unlikely to explain much of the variation in the flux depression measurements.

The probable source of the wide range observed (around a factor of 2 at $z = 3$) is variations in the background UV flux levels related to the local density of radiation sources.

4. Flux Statistics and $J_{-21}$

The transmitted flux, $F = e^{-\tau}$ in a spectrum is a measure of the opacity of the medium. It has advantages over line fitting only in that it is completely objective. The flux decrement is $D = 1 - F$. $D \sim \tau$ if $\tau$ is small. $D_A = < D >$ is the mean flux decrement due to HI Lyα absorption. It is measured between 1026 Å (Lyman-β) and
1216 Å in the quasar rest frame where contamination from other lines is minimal.

We calculate flux statistics such as the flux decrement one-point distribution directly from the artificial spectra. We can then determine the ionizing radiation levels required for the simulations to match the observed cumulative flux decrement distribution: \( C(D) \). The scaled distributions match the observed distribution well (figure 4.10).

Matching the cumulative flux decrement distribution gives a very similar result to matching the observed mean flux decrement \( D_A = \langle D \rangle \) because the integral of the cumulative distribution is the mean flux decrement, as shown in the following equations:

\[
C(D) = \int_{-\infty}^{D} P(D')dD', \\
D_A = \int DP(D)dD, \\
= \lim_{X \to -\infty} X - \int_{-\infty}^{X} (1 - C(D))dD
\]  

(4.1)

(4.2)

\( P(D) \) refers to the likelihood of having flux decrement \( D = 1 - e^{-r} \) at a random point in the spectrum. The integrals ideally only go from 0 to 1, but the normalized spectra spill outside these bounds due to noise. The shape of the distribution is most sensitive to the mean flux decrement desired and not the redshift of the data so that the redshift 2 distribution may be rescaled to fit the redshift 3 distribution quite well. The discrepancies shown in figure 4.10 are located near zero and maximum flux and related to noise treatment. We modelled the noise very simply. Matching these portions of the distribution require very detailed noise modelling (Rauch et al. (1997b)).

We re-processed our simulations to produce new artificial spectra with values of the ionizing radiation background chosen to fit the observed flux decrement distributions given in Rauch et al. The values of \( J_{-21} \), the flux at Lyman edge \( (J_{\nu} = J_{-21} 10^{-21} \text{erg s}^{-1} \text{Hz}^{-1} \text{cm}^{-2} \text{Sr}^{-1}) \) for each of the various cosmologies and normalizations studied are given in table 3. Fits to the cumulative flux density
distribution curves are shown in figure 4.11. The fits at $z = 2$ are acceptable considering the intrinsic variation in the data is at least 0.05 vertically on the figure. All four cosmologies with similar rms fluctuation amplitudes filtered at 0.5 Mpc ($\sigma_{G,0.5} = 1.05$) fit the data. The two SCDM simulations with $\sigma_8 = 0.44$ and 1.0 are highly inconsistent with the data.

Normalizing to the Rauch et al. flux decrement distribution does not produce a good fit to the column density distribution. As stated in Chapter one, there is a large amount of variation in flux decrement measurements among lines of sight. The $N_{HI}$ column density distribution comes from a larger data set consisting of different quasars to the seven studied by Rauch et al.. The flux decrements determined by Zuo & Lu (1993) are consistent with our initial choice of ionizing flux level $J_{-21} = 0.5$ for $z = 2 - 3$. However, we expect that the typical mean flux decrement is closer to the value inferred from the high quality data of Rauch et al.

Table 4 shows the flux decrements resulting if $J_{-21} = 0.5$ is used at all redshifts. The simulations marked with an asterisk have the same small scale fluctuation amplitude $\sigma_{G,0.5} = 1.05$ at $z = 3$. A large contributor to the high flux decrements for $\Lambda$CDM is the relatively low velocities in that cosmology at high redshifts. There is thus relatively more material per unit redshift and thus per unit wavelength to produce opacity. The conversion from physical size to velocity-space size is made

<table>
<thead>
<tr>
<th>Cosmology</th>
<th>$\sigma_8$</th>
<th>$J_{-21}$ (z=2)</th>
<th>(z=3)</th>
<th>(z=4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCDM</td>
<td>1.0</td>
<td>0.14</td>
<td>0.12</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>0.67</td>
<td>0.24</td>
<td>0.19</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>0.44</td>
<td>0.29</td>
<td>0.23</td>
<td>0.19</td>
</tr>
<tr>
<td>$\Lambda$CDM</td>
<td>0.93</td>
<td>0.30</td>
<td>0.19</td>
<td>0.12</td>
</tr>
<tr>
<td>OCDM</td>
<td>0.81</td>
<td>0.22</td>
<td>0.16</td>
<td>0.13</td>
</tr>
<tr>
<td>HCDM</td>
<td>0.82</td>
<td>0.24</td>
<td>0.18</td>
<td>0.13</td>
</tr>
</tbody>
</table>

Table 3: UV background required to match observed flux decrement distribution.
Table 4: Measured $D_A$. $J_{-21} = 0.5$ for the simulations.

<table>
<thead>
<tr>
<th>Redshift</th>
<th>$Z=2$</th>
<th>$Z=3$</th>
<th>$Z=4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rauch</td>
<td>0.148</td>
<td>0.316</td>
<td>0.543</td>
</tr>
<tr>
<td>Press</td>
<td>0.15</td>
<td>0.36</td>
<td>0.62</td>
</tr>
<tr>
<td>Schneider</td>
<td>0.2</td>
<td>0.35</td>
<td>0.55</td>
</tr>
<tr>
<td>Zuo &amp; Lu</td>
<td>0.08</td>
<td>0.22</td>
<td>0.63</td>
</tr>
<tr>
<td>Simulation $\sigma_8$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCDM</td>
<td>1.0</td>
<td>0.065</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>0.67*</td>
<td>0.10</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>0.44</td>
<td>0.11</td>
<td>0.22</td>
</tr>
<tr>
<td>SCDM LR</td>
<td>0.67</td>
<td>0.056</td>
<td>0.14</td>
</tr>
<tr>
<td>LCDM</td>
<td>0.93*</td>
<td>0.13</td>
<td>0.24</td>
</tr>
<tr>
<td>OCDM</td>
<td>0.81*</td>
<td>0.10</td>
<td>0.21</td>
</tr>
<tr>
<td>HCDM</td>
<td>0.82*</td>
<td>0.11</td>
<td>0.22</td>
</tr>
</tbody>
</table>

using the Hubble parameter:

$$H(z) = H_0 (\Omega_{nr}(1 + z)^3 + \Omega_{curv}(1 + z)^2 + \Omega_\Lambda)^{1/2}. \quad (4.3)$$

where $\Omega_{nr} = \Omega_{dm} + \Omega_b$ is non-relativistic matter. $\Omega_{curv}$ is the term representing curvature, if present, and $\Omega_\Lambda$ is the vacuum energy-density at $z = 0$. The mean opacity is proportional to $(\Omega_b h^2)/J_{91_2}/H(z)$, making the assumption that the neutral hydrogen density goes as the square of the baryon density. (e.g. at $z = 3$. $H_{SCDM, HCDM} = 400 \text{ km s}^{-1}$, $H_{LCDM} = 330 \text{ km s}^{-1}$ and $H_{OCDM} = 406 \text{ km s}^{-1}$). If we compensate for this factor the differences between the cosmologies become much less.

The large difference in the flux decrements for the high and low resolution SCDM simulations can be explained in terms of the gas that provides most of the opacity. In figure 4.12 we show the probability of different densities in our simulations. Improving resolution results in more tight gaseous lumps and emptier voids. This is naturally to be expected as small scale structures form first. If we lower the resolution to exclude these waves, the medium is closer to a uniform mean
density.

The mean flux decrement is dominated by lines below $N_{HI} \sim 10^{14}$ cm$^{-2}$ at $z \sim 3$ as shown in Chapter one. These lines are generated by gas with overdensities around 10, depending on redshift. There is much more gas at these densities in the higher resolution simulations, as shown in the figure. In the optically thin limit, where $e^{-\tau} \sim 1 - \tau$, the flux decrement $D_A$ is roughly proportional to the density in neutral hydrogen:

$$D_A \sim <\tau> \propto <N_{HI}> = \frac{1}{L} \int n_{HI} dl.$$  \hspace{1cm} (4.4)

$<\rho^2>$ is higher for our high resolution at all redshifts and $<n_{HI}> \propto <\rho^2>$ is correspondingly greater. Much of this gas is in mildly overdense, filamentary structures where it provides a lot of opacity. We deliberately exclude very high density gas from the averages discussed for this argument because very dense structures contribute little to the mean opacity at $z \gtrsim 2$.

The increase in the opacity associated with increasing the typical fluctuation amplitude may be expected to occur as $\sigma_8$ is increased for the three high resolution SCDM cases. When the fluctuation start to collapse, however, the gas collects in objects which produce saturated lines. The reasoning of 4.4 then fails because the opacity due to saturated lines is minimal. The $\sigma_8 = 1.0$ case always suffers from this compared to the $\sigma_8 = 0.67$ case. However, the $\sigma_8 = 0.44$ case lacks structure at $z = 4$ and thus has a lower opacity when compared to the $\sigma_8 = 0.67$ case there. As time progresses, gas in the $\sigma_8 = 0.67$ case collects in objects and its opacity becomes lower than that of the $\sigma_8 = 0.44$ case by $z = 2$. 


Fig. 4.7.— Several sample lines of sight through the $\nu = 0$ SCDM simulation at $z = 3$. In each panel, the spectrum is shown, with random pixel noise added and the fits performed by the automated Voigt profile fitting routine overlaid in grey.
Fig. 4.8.— Four lines of sight through the $\nu = 0$ CDM simulation at $z = 3$. The lowest panels are the spectra, with the fits performed by the automated Voigt profile fitting routine. Above these we show the mean gas temperature, weighted by neutral hydrogen. The second from top panels show the actual velocity space distribution of neutral gas. The top panels are contour maps (in physical space vertically and velocity space horizontally) of nearby neutral gas – showing the nearby structures mappable with unrealistically densely packed lines of sight. The main line of sight is drawn in.
Fig. 4.9.— The same two lines of sight through the ν = 0 simulation of each cosmology at \( z = 3 \). The cosmologies, clockwise from the top left, are SCDM, HCDM, OCDM and ΛCDM.
Fig. 4.10.— Scaling of the cumulative flux decrement distribution. The small differences between the observed data at $z = 3$ (from Rauch et al. (1997b)) and the simulations is not significant and arises mostly due to the different structure of the noise and the point spread function in the observed spectra, which we did not model in detail. The dashed line are the simulated data before rescaling.
Fig. 4.11.— Scaling of the cumulative flux decrement distribution at three redshifts. In each panel, the thick lighter lines represent data for redshifts $z = 2$, $3$ and $4$ from Rauch et al. (1997b) going from left to right, respectively. The solid lines are simulated results with rescaled UV and the dashed lines are for $J_{21} = 0.5$. The first four panels show four cosmologies: upper left SCDM ($\sigma_8 = 0.67$), upper right HCDM, middle left LCDM and middle right OCDM. The lower left and right panels are SCDM with $\sigma_8 = 0.44$ and $1.0$ respectively.
Fig. 4.12.— The probability distribution for gas density. As expected, with higher mass resolution (100 kpc) the voids are more empty and more high density gas is present (including gas in low mass objects missed at low resolution) than at low resolution (312.5 kpc). The models shown here are standard CDM with $\nu = 0$. The results presented in Rauch et al. (1997b) are a low resolution SCDM (slightly lower peaked curve) and an Eulerian $\Lambda$CDM model (both shown dotted).
4.1. Visual Comparison among Cosmologies

We now make a visual comparison of the way the cosmologies look in neutral hydrogen. Figure 4.13 shows the simulations with the standard value of $J_{-21}$ used. The OCDM and $\Lambda$CDM especially have less small scale power than the SCDM case and this is apparent in the way the filaments are more coherent. The small scale power has acted to bead up the filaments more in the SCDM ($\sigma_8 = 0.67$) case. In figure 4.14 the ionizing flux level has been lowered to better fit the flux decrements as discussed in the previous section. The main feature to note is the increased prominence of the filaments.

Figure 4.15 shows both fixed and adjusted ionizing flux $N_{HI}$ column densities for the high and low $\sigma_8$ SCDM cases. Note the massive drainage of voids and filaments for the $\sigma_8 = 1.0$ case shown in the top left panel. Alternately, for $\sigma_8 = 0.44$ there is a lot of material in the voids. After the ionizing flux is scaled the $\sigma_8 = 1.0$ filaments are fatter (lower density gas contributes significant absorption) and there are abundant high column absorbers.

5. Voigt Profile Fits

We fit Voigt profiles using an automated profile fitting program, designed to emulate the methods employed by observers. An automated program was used for speed and to maintain a consistent set of line selection criteria for all simulations and redshifts. Several thousand lines are fit per axial direction for a given simulation. The statistically complete $N_{HI}$ frequency curve for each cosmology at $z = 4$ required the fitting of 300,000 lines.

The program identifies a line group region as a portion of the spectrum where the flux drops below 98% of the continuum. The procedure to fit lines in each group region is as follows:

1. (a) Locate a first line centre at the lowest point in the line group region of the
Fig. 4.13.— $N_{\text{HI}}$ Column density through 2 Mpc thick slices at $z = 3$. The cosmologies shown, clock-wise from the top left, are SCDM, HCDM, OCDM and $\Lambda$CDM. The ionizing flux used to create the artificial spectra was the value used during the simulations: $J_{-21} = 0.5$. The contours in $\log_{10} N_{\text{HI}}$ are 11-12.75 (dark), 13 (dark grey), 13.5 (medium grey), 14 (light grey), 14.5 (dark), 15 (black) and 17 (white).
Fig. 4.14.— $N_{HI}$ Column density through 2 Mpc thick slices at $z = 3$. The cosmologies shown, clock-wise from the top left are SCDM, HCDM, OCDM and LCDM. $J_{-21}$ has been adjusted in each case to fit the flux decrement observations of Rauch et al. (1997). The contours in $\log_{10} N_{HI}$ are 11-12.75 (dark), 13 (dark grey), 13.5 (medium grey), 14 (light grey), 14.5 (dark), 15 (black) and 17 (white).
Fig. 4.15.— $N_{HI}$ Column density through 2 Mpc thick slices at $z = 3$. We show the effect of adjusting $\sigma_8$ on the $\nu = 0$ SCDM initial conditions. The top panels are for a fixed ionizing flux level of $J_{-21} = 0.5$. In the lower panels, $J_{-21}$ has been adjusted as shown to fit the flux decrement observations of Rauch et al. (1997). The contours in $\log_{10} N_{HI}$ are 11-12.75 (dark), 13 (dark grey), 13.5 (medium grey), 14 (light grey), 14.5 (dark), 15 (black) and 17 (white).
spectrum.

(b) Fit a line by $\chi^2$ minimization localized around the minimum (not the whole group).

(c) Attempt to fit the whole line group region by $\chi^2$ minimization with one line using this line as a first guess.

(d) Calculate the reduced-$\chi^2$ of this fit (see below).

2. (a) Locate another line centre at the lowest point in the line group region of the spectrum with absorption due to previous lines from Step 1b) removed.

(b) Fit the line by $\chi^2$ minimization localized around the minimum.

(c) Attempt to fit the whole line group region by $\chi^2$ minimization with the current lines as above (allowed to vary in all three parameters)

(d) Calculate the reduced-$\chi^2$ of this fit.

3. Return to step 2. until two iterations have failed to produce a better fit (lowered reduced-$\chi^2$).

The reduced-$\chi^2$ is given by $\chi^2/d$. $d$ is the number of degrees of freedom remaining for the fit. estimated as 1 per 2 pixels ($10$ km s$^{-1}$) in the line group region minus 3 for each line used in the fit. This inhibits the fitter from producing fits with more lines than are warranted by the number of data points. In practice a line group rarely needs more than 6 lines for a good fit. We employed the Levenberg-Marquardt routine from Press et al. (1992). This algorithm efficiently finds local minima in functions of many variables. Occasionally it blows up or stalls in non-global minima. We added local Monte Carlo type tests for smaller values of $\chi^2$ short distances away in the parameter space. Monte Carlo is slow in itself but using it periodically keeps the high order Levenberg-Marquardt algorithm from getting into a rut.

To validate our fitting technique, we made use of Keck data provided by Kirkman and Tytler (1997) and compared our fits to theirs. A comparison is shown in figure 4.16. Our line lists are not identical but that is expected as there is not an ideal fit. In particular, at columns around $10^{16}$ cm$^{-2}$ it is very hard to break the degeneracy between $N_{HI}$ and the line broadening, $b$, because the damping wings
Fig. 4.16.— Performance of Automated Voigt fitter on real data. The data used is from Kirkman and Tytler (1997). The discrepancies present are similar to the normal variation expected in line fitting – there being no single perfect fit.
are not yet well established – thus \( N_{HI} \) is hard to determine without additional information, e.g. Lyman-3. The Voigt profile fitting method gives great physical insight into the gas structures that form the observed spectra. The most obvious drawback of the technique is its subjective nature. Tailoring the Voigt fitting routine to select lines in a similar fashion to the line list in Kirkman and Tytler (1997) generates substantial differences in the column density distribution. Kirkman and Tytler's fit has a break in the power law at \( 10^{13.5} \text{cm}^{-2} \) (figure 1.4), where as the larger data set of Petitjean et al. (1993) combined with the data of Hu et al. (1995) for low columns has a break at \( 10^{14.5} \text{cm}^{-2} \) which is reproduced by the original unmodified version of the automated fitter (see figures 4.17 and 4.20). This result is not entirely due to the subjective nature of fitting, but also a resolution effect as the Kirkman and Tytler (1997) data set has resolution (FWHM) of 8 km s\(^{-1}\) against \( \sim 25 \text{ km s}^{-1} \) for the Petitjean et al. (1993) data. We will make our comparisons against the Petitjean et al. (1993) data as it is the more extensive set and has been available for longer. General conclusions can be drawn independently of the exact positioning of the break feature.

Our standard fitter was not designed to suit to any particular data set or line list but to appropriately fit the lines in our simulated spectra according to our own criteria, with reference to our exact knowledge of the underlying gas velocity space density distribution. All these fits are "best fits" in the \( \chi \)-square minimized sense. Ideally, one would employ a method that was purely objectively defined in place of Voigt fits but the information gained through Voigt fitting is more open to direct interpretation than alternate methods suggested, though 2-point measures look promising (e.g. Miralda-Escudé et al. (1997)).

We can use this simulation information to validate fitting in general. In the figures in section 2. we included comparisons our Voigt fits to characteristic objects along lines of sight – showing how accurately the columns of neutral material have been measured.
Fig. 4.17.— $N_{HI}$ Column density distribution for the simulation sample comparing two versions of the fitter.
6. Properties of Gas Associated with Lines

For each line fit for our SCDM simulation, we sampled the properties of the gas at the line centre and the results are shown in figure 4.18. The equilibrium curve shown for the temperature includes the effect of the general adiabatic expansion but not local compression or shock heating. Shock heating is largely responsible for the additional temperature above the equilibrium curve at moderate to high densities or column densities. At low densities, the gas is not in temperature equilibrium. The cooling times are long so it retains the higher temperatures it had at earlier times when it was denser.

There is a very strong relationship between the column density and the gas density at the line centre. The numerous lines around \(N_{HI} \sim 10^{13}\) cm\(^{-2}\) are strongly associated with mean density gas. The points scattered above the typical range, given by the diamonds and vertical bars, are associated with small lines superimposed upon a larger line as a blend. The gas density measurement at line centre corresponds to that for the higher column line. This effect is not easy to remove and for this reason the points are upper limits. A 0.1 Mpc thick slab with mean density at \(z = 3\), \(J_{-21} = 0.5\) and \(T = 20,000 K\) would give a column density of \(N_{HI} \sim 10^{12.3}\) cm\(^{-2}\). This is consistent with the points in the bottom right panel in the figure.

7. Column Density Distribution

The column density distribution is often plotted as lines per unit interval in a variable, \(X = 0.5((1 + z)^2 - 1)\), where \(z\) the redshift (Petitjean et al. (1993)). This variable is used as a “radial coordinate”. It was apparently designed to be appropriate to open universes, which is incorrect. We retain the use of \(X\) because it is how the observational data are plotted. It affects only the normalization of the curve and does not confuse the results. To convert to lines per unit redshift interval,
Fig. 4.18.— Temperatures and baryon densities associated with lines at \( z = 2 \) and \( z = 3 \). This can be compared with figure 4.6 from the particles. The blue diamonds represent the median values and the vertical bars span the first to third quartile. The lowest panels have density divided by mean density shown.
Fig. 4.19.— Column density distribution for our total simulations sample, \( z = 3 \). The plot covers 10 orders of magnitude and compares the SCDM result to the data of Petitjean et al. (1993) (black boxes) and Hu et al. (1995) (dotted and dashed boxes) \((z_{\text{obs}} \sim 2.7)\). The boxes represent bins with the top and bottom of each box being the Poisson error. The dotted Hu et al. boxes have had a blending correction applied which has not been attempted with the simulated lines. We have adjusted the UV flux in order to fit the observed flux decrement 1 point distribution. The simulation fails to reproduce numbers of clouds above \(10^{16}\) correctly because of self-shielding of the ionizing flux at high columns.
\[ \frac{df}{(dN_H/dz)} \], we would multiply the \( \frac{df}{(dN_H/dX)} \) data by \((1 + z)\). This would require knowledge of the exact redshifts of the observations, which are actually typically spread over a range of \( \sim 0.5 \) in redshift for each spectrum. The evolution of lines above a column density cut goes roughly as \( dN/dz \propto (1 + z)^\gamma \), \( \gamma \sim 2.5 \) (though the power law varies depending on redshift, dropping drastically below \( z \sim 1 \) (see Chapter One)). Thus plotting versus unit \( X \) partially removes the change in normalization with redshift. A better variable with which to compare shape changes in the spectrum for data at different redshifts would be \( Y = \int (1 + z)^\gamma dz \), for which no normalization changes occur (as used in proximity effect studies).

The physical differential distance per redshift interval is given by.

\[
\frac{dl}{dz} = \frac{1}{(1 + z)H(z)} = \left[ (1 + z)H_0(\Omega_{\text{nr}}(1 + z)^3 + \Omega_{\text{curv}}(1 + z)^2 + \Omega_\Lambda)^{1/2} \right]^{-1}
\]

where \( H(z) \) is the hubble parameter \(-\dot{z}/(1 + z)\). and \( \Omega \) refers to the density in a species as a fraction of the critical density at redshift \( z = 0 \). \( \rho_{\text{crit}} = 3H_0^2/(8\pi G) \). \( \Omega_{\text{nr}} = \Omega_{\text{dm}} + \Omega_\Lambda \) represents non-relativistic matter. \( \Omega_{\text{curv}} \) is the term representing curvature and \( \Omega_\Lambda \) is the vacuum energy-density.

Numerical simulation output is in terms of velocity which we must convert to redshift. The conversions between redshift, wavelength and velocity are as follows.

\[
1 + z = \frac{\lambda_{\text{obs}}}{\lambda_{\text{rest}}} \quad (4.7)
\]
\[
\Delta z = (1 + z)\Delta v_{\text{local}}/c \quad (4.8)
\]
\[
\Delta \lambda_{\text{obs}} = \lambda_{\text{rest}}(1 + z)\Delta v_{\text{local}}/c \quad (4.9)
\]

Figure 4.19 is a standard plot of the column density distribution, with our standard CDM result plotted. The plot is not very informative because of the 10 orders of magnitude it covers. We strongly advocate removing a power-law similar to \( 10^{8.25}N_H^{-1.46} \). We chose this powerlaw to match the fit of Hu et al. (1995) to the
low column end of the distribution. This subtraction reveals the detail present in the distribution. All following plots will have the power law subtracted.

Figure 4.20 clearly demonstrates the agreement between our results and the observations. The curve with the flux adjusted is slightly to the right of the observational results – due to the combined effects of the small difference in the redshifts of the observations and the simulated data and the slight inconsistency between the flux decrement results of Rauch et al. (1997b) and the column data. Looking at the results for decreasing redshift (figure 4.21) the progression on the break in the power law is clearly evidenced. This progression of this break has been observed (Kim et al. (1997), explicitly compared in figure 4.22) as a deficit of intermediate column lines at lower redshifts. This progression is a strong prediction of our work. This evolutionary trend can be understood in terms of the fact that the structures responsible for absorption evolve slowly above redshift 2, staying fairly static in comoving space. The overwhelming factors are the drop in density with the cosmic expansion and the evolution of the UV flux. Given a system of cosmologically expanding clouds, we can crudely estimate how the column of one cloud evolves as follows.

\[ n_{H1} \propto n_b^2/J_{-21}(z), \quad (4.10) \]

\[
N_{H1,\text{cloud}} = \int n_{H1}(r) \, dr \\
\propto \int n_b(x)^2/J_{-21}(z) \, d(x/(1+z)) \\
= (1+z)^5/J_{-21}(z) \int n_{b,z=0}(x)^2 \, dx. \quad (4.11)
\]

where physical distance \( r = x/(1+z) \). The relation between \( n_b \), the baryon density, and \( n_{H1} \), the neutral hydrogen density, holds for \( n_b << 1 \) which is true everywhere in the medium except for the cores of dense clumps. Given the ionizing flux values in table 3, equation 4.11 provides a good fit to the simulated data. We expect the break to move roughly 0.7 in \( \log_{10} \) from \( z=2 \) to 3 and similarly from \( z=3 \) to 4. Evolution in the structures goes in the other direction, increasing the density. Figure 4.3 is
Fig. 4.20.— Column density distribution at $z = 3$. The plots compare the overall results for four cosmological models to the data of Petitjean et al. (1993) (black boxes) and Hu et al. (1995) (dotted and dashed boxes) ($z_{\text{obs}} \sim 2.7$). The boxes represent bins with the top and bottom of each box being the Poisson error. The dotted Hu et al. boxes have had a blending correction applied which has not been attempted with the simulated lines. We have subtracted the power law $-1.46 \log_{10}(N_{HI}) + 8.25$. In the left panel $J_{-21} = 0.5$ and in the right we have adjusted the UV flux in order to fit the observed flux decrement one-point distribution. Note the features in the distribution, including the hump at $\sim 10^{14}$ indicating a transition between low overdensity material and high density material near dwarf galaxies. The dynamics and physical properties of the gas change markedly above this column density. The simulation fails to reproduce numbers of clouds above $10^{16}$ correctly because of self-shielding of the ultraviolet flux at high columns.
a visual demonstration of the way in which the structure tends to remain fixed in comoving space.

An clear result apparent in figure 4.20 is the lack of strong signatures of the underlying cosmology. This can be explained in terms of the freedom to renormalize the curve to fit the data. If the UV flux was better constrained we might be able to use it as a cosmological constraint. One current estimate for the UV flux level from the proximity effect. the depression in absorber counts near the QSO source. is \( e.g. \)

\[
J_{912A} = 5^{+2.5}_{-1} \times 10^{-22} \text{erg cm}^{-2} \text{s}^{-1} \text{Hz}^{-1}
\]

with no evolution over the range \( z = 1.7 - 4 \) Giallongo et al. (1997). Our simulations indicate a need for a factor of 2 increase in the flux between redshift \( z = 4 \) and \( z = 2 \). The uncertainty in \( \Omega_b \) is large but it’s value must be fixed at all redshifts. The evolution we predict in the flux is consistent with the predictions from the proximity effect studies – which must contend with a lot of uncertainty. Current knowledge of the evolution of the number density of quasars also indicates an evolution of order of a factor of 2 in the ionizing flux over this range (Haardt & Madau (1996)).

Section 9. investigates the source of the break in the the power-law with the conclusion that it is purely a function of the gas density responsible for the absorbers of a given \( N_{HI} \) column. The existence of the break in the column density distribution is very important because it means the distribution is not invariant under the ionizing flux scaling. The break is a smooth feature and it does not move much as the cosmology is changed. If simulations were performed and the resulting artificial spectra fit exactly in the same way as a reasonably large absorber data set then we could attempt to discriminate among cosmologies using the break position. With the quality of the current data and the uncertainties related to the fitting procedure, we can not be so optimistic. It should also be possible to remove the discrepancy between the column density results and the flux decrement results, at least on a quasar-by-quasar basis.

The column density distribution is essentially a one point function – and it is hard to extract information about the relative power on different scales. Such
information is usually derived from correlations. Measurements of correlation in the Lyman-α forest (e.g. Hu et al. (1995)) are difficult to compare with simulations because the important scales are large compared to the simulation boxes. For periodic boxes, structure on roughly half the box size is already strongly correlated due to the periodicity. The peak patch approach discussed in Chapter Three is an alternate to direct $N$-body/ Hydrodynamical numerical simulations that accurately includes large scale correlations.

Evolution in the column density distribution is only just beginning to be observed (Kim et al. (1997)). Previously it had often been stated in the literature that no evolution is consistent with the column density data (Kirkman and Tytler (1997). Petitjean et al. (1993)). Our simulations imply strong evolution in the column density distribution consistent with the Kim et al. results. In section 12. we undertake to separate out the structural elements responsible for absorption features. This promising avenue should result in a stronger statement on evolution by tying specific absorption line columns to specific structures.

We also find no strong signature of $\sigma_8$ in the column density distribution after the flux is rescaled. Figure 4.23 shows the result with and without UV flux rescaling to fit the flux decrement distribution. After rescaling the $\sigma_8 = 0.44$ results is slightly different but the $\sigma_8 = 0.67$ and $1.0$ results are very similar. This reinforces the important result that the rescaling possible undermines the effectiveness of fits to the column density for discriminating between cosmologies. The uncertainty present in the the UV flux level combined with that in $\Omega_b$ is sufficient to allow the rescaling of all but the most extreme results onto the observational data.
Fig. 4.21.— Redshift evolution of the Column density distribution for the four cosmologies. The top, middle and bottom panels show $z = 2$, 3 and 4 respectively. In the left panels the flux is $J_{-21} = 0.5$ and in the right we have adjusted the UV flux in order to fit the observed flux decrement one-point distribution at each redshift. Note the progression of the features in the distribution, especially the break, to lower column at lower redshifts.
Fig. 4.22.— The Column density distribution evolution. The data is from Kim et al. (1997). The lower four simulation lines are for $z = 2$ with $J_{-21} = 0.5$ and the upper four are for $z = 3$ with $J_{-21} = 0.2$. The break in the power law at roughly $10^{14}$ moves to lower columns, as a result of decreasing overall density with lowering redshift.
Fig. 4.23. — \( \sigma_8 \) Variation. The Column density distribution at \( z = 2, 3 \) and 4 (going from top to bottom). In the left panels the UV flux is fixed at \( J_{-21} = 0.5 \) and in the right the flux has been adjusted to fit the observed flux decrement one-point distribution.
8. Line Widths

Figure 4.24 shows the line width distribution from the raw simulation output computed with the fixed ionizing flux level used during the calculations ($J_{-21} = 0.5$). The simulated results are the weighted sum of 5 sets of $\sim 3000$ lines with column densities $N_{HI} = 3 \times 10^{13} - 3 \times 10^{14} \text{ cm}^{-2}$. The evolution is to the right with decreasing redshift as the systems become physically and dynamically hotter. There is no great difference between the results for different cosmologies. The lines beyond the last bin have been added to the last bin. Comparing the dotted (thermal only line width) and solid (measured line width) shows that the effect of velocity broadening is to add $\sim 10 \text{ km s}^{-1}$ to the line widths.

Figure 4.25 shows the simulation line width distributions with scaled ionizing flux levels to match the HI Lyman-$\alpha$ flux decrement given in Rauch et al. (1997b). The rescaling required is given in table 3. Overlaid is data from Kim et al. (1997). The two bins of Kim et al. observations include $\sim 200$ lines each. The rescaling procedure changes the neutral fraction by a fairly uniform multiplier across the simulation. The $\Lambda$CDM case required the least rescaling. The rescaling implies that the gas responsible for the absorption in the fixed column range is lower density compared to that for the $J_{-21} = 0.5$ case. For this reason the temperatures are colder. If the lowered flux had been used during the calculation, the temperatures in the simulation itself would also be slightly smaller but this is not expected to be a large effect (maximum potential change in line widths $\sim 2 \text{ km s}^{-1}$ and only at high columns, see Chapter 2). Results are similar for all cosmologies. A strong degree of similarity is expected as the rescaled ionizing flux levels are similar for each cosmology.

In figure 4.27, the median line widths for 4 cosmologies are shown. Shown in grey are the results for the individual calculations with different background field levels ($\nu$). The grey labels from -1.1 or 1.4 indicate the ordering vertically. The results are for lines in the column ranges $3 \times 10^{13}$ to $3 \times 10^{14} \text{ cm}^{-2}$. There are
Fig. 4.24.—Line width $b$ distributions for 4 cosmologies and fixed ultraviolet flux levels $J_{21} = 0.5$. The panels, clockwise from the top left, are for SCDM, HCDM, and OCDM respectively. The dashed curves are the distributions for the thermal part of the broadening for these lines and $z = 2$ respectively. The dashed curves are the distributions for the thermal part of the broadening for these lines.
Fig. 4.25.— Line width distributions for 4 cosmologies and rescaled ultraviolet flux levels to match the HI Lyman-α flux decrement. The panels, clockwise from the top left are for SCDM, HCDM, ΛCDM and OCDM respectively. The top part of each panel has the $z = 2$ and $z = 3$ simulation results overlaid with a histogram of the Kim et al. data for $<z> = 2.31$. The lower part has the $z = 3$ and $z = 4$ simulation results overlaid with Kim data for $<z> = 3.35$. 
Fig. 4.26.— Line width $b$ distributions for differing fluctuation level normalizations and fixed ultraviolet flux levels $J_{-21} = 0.5$. The upper and lower panels are for the SCCM runs with $\sigma_8 = 0.44$ and 1.0 respectively. In the left panels the curves from left to right refer to $z = 4$, $z = 3$ and $z = 2$ respectively. The dashed curves are the distributions for the thermal part of the broadening for these lines. The right panels are as in the previous figure.
order of 3000 lines in the total sample from one simulation. The simulations have been statistically combined and the results are indicated with a plus symbol at the relevant redshift. The data from Hu et al. (1995) is for column ranges $10^{13.1} \text{ cm}^{-2}$ to $10^{14} \text{ cm}^{-2}$ and $10^{13.8} \text{ cm}^{-2}$ to $10^{16} \text{ cm}^{-2}$ for the dotted and solid lines respectively. The quasars responsible are given by the labels. The steep slope of the column density distribution means that lines near the lower edge will dominate the samples. The simulated results could partially explain the variability in the observations.

A lack of peak-like regions (similar to $\nu = 1.4$) in the line of sight would depress the values substantially. Local variation in the ultraviolet background, while not changing the equilibrium temperatures will change the density/column density correspondence and thus give different temperatures.

The two main contributors to line width are temperature and velocities. The two are closely related and indicate increased velocity dispersion in dark matter halos for higher temperatures. The expanding voids, $\nu < 0$, can exhibit greater velocities than the mean $\nu = 0$ case at early times, but structure formation quickly freezes out and at late times the voids exhibit smaller line widths. The peaks form collapsed objects with large velocity dispersions and undergo dramatic shock heating at late times. This violent process generates the large line widths shown in the figures for $\nu = 1.4$.

Figure 4.27 shows the SCDM ($\sigma_8 = 0.67$) to be more dynamic than the other cosmologies at $z = 2$. Aside from possessing more small scale power, it also undergoes the most linear growth after $z = 3$ when $\sigma_{G,0.5} = 1.05$ for all four cases. Figure 4.28 includes the UV flux rescaling necessary to fit the flux decrements. This lowers the line widths because the lowered ionizing flux promotes lower density, cooler gas absorbers to the column range selected.

In figure 4.29 we demonstrate the effect of varying $\sigma_8$ for the standard cold dark matter case. The high-$\sigma_8$ case is much more dynamic, but draining of gas into objects has required a lower ionizing flux to fit the HI Lyman-\(\alpha\) flux decrement. These greater dynamics and the lower flux pull in opposite directions. The high-$\sigma_8$
Fig. 4.27. — Median b values for a fixed ultraviolet flux level of $J_{-21} = 0.5$ (used during the simulations). The labelled observational data is from Hu et al. (1995). The panels, clockwise from the top left are for SCDM, HCDM, ACDM and OCDM respectively. The comparison should be made between the dotted data and the crosses. Details are given in the text.
Fig. 4.28. — Median line width values for ultraviolet flux levels varied below the values used during the simulations to match the opacity result of Rauch et al. (1997b). The lowered flux levels would lower the values by as much as 2 km s$^{-1}$ for a fully consistent simulation result. $J_{-21} = 0.5$ was used during the simulations. The panels, clockwise from the top left, are for SCDM, HCDM, ACDM and OCDM respectively. The labelled observational data is from Hu et al. (1995).
case absorbers arise in lower density gas which is colder than the higher density gas responsible for the low-$\sigma_8$ absorbers. The dynamics are thus masked after ionizing flux renormalization.

The low resolution run line-width results, presented in figure 4.30 are much higher for the same fixed flux level. The cosmology and other parameters are identical to the high resolution standard CDM runs. Linking this result with the vast reduction in line numbers (demonstrated in Chapter Two), the higher line widths are mostly likely the result of the lines arising in higher velocity dispersion objects and large structures with a large velocity variation across them. High resolution runs also produce these structures but more smaller scale structures. The typical lines at these columns ($\sim 10^{14}$ cm$^{-2}$) are contributed by small filaments around around objects that are not resolved in the low resolution runs. This difference is shown in figure 4.31. There are roughly half as many lines in this column range at low resolution. The figure suggests that the same objects could contribute the high line width absorbers in both cases. An exact correspondence has not been determined, though we have studied objects in both cases and determined that low mass objects and small scale structure we know to be responsible for many lines are absent at lowered resolution. The simulations use different initial conditions so a direct object-by-object comparison is not possible.
Fig. 4.29.— $\sigma_8$ variation. Median line width values for a fixed ultraviolet flux level ($J_{-21} = 0.5$) on the left and fluxes varied below the values used during the simulations to match the opacity result of Rauch et al. (1997b) on the right. The upper and lower panels are standard CDM simulation results with $\sigma_8=0.44$ and 1.0 respectively. The labelled observational data is from Hu et al. (1995).
Fig. 4.30.— The median line width results for the low resolution runs. The results are much higher at all redshifts despite the ultraviolet flux level being the same ($J_{\text{-21}} = 0.5$). The labelled observational data is from Hu et al. (1995).
Fig. 4.31.— The line width distributions for $N_{HI}$ columns $3 \times 10^{13}$ to $3 \times 10^{14}$ cm$^{-2}$. The dotted lines are the high resolution standard cold dark matter results. The solid lines are the result at three times lower resolution. The results from left to right are for redshifts $z = 2$, 3 and 4 respectively. The area under the curves in the two panels is proportional to the numbers of lines. There are roughly half as many lines in this column range at low resolution.
9. Rescaling

There are two kinds of rescaling which are performed on simulations: the values of $\Omega_b$ or $J_{-21}$ may be changed after the fact. For the Lyman-$\alpha$ forest conditions it is commonly assumed that the opacity scales as $\Omega_b^2 h^4/(J_{-21} H(z))$ (e.g. Rauch et al. (1997b)). We show that this equation lacks an highly significant term – temperature. The recombination rate is a sensitive function of temperature ($r_{\text{rec}} \propto T^{-\alpha_{\text{rec}}}$, $\alpha_{\text{rec}} \sim 0.7$), so if the processing has these parameters set at values sufficiently different to those used during the simulation the conclusions will be incorrect. Fortunately, temperature is highly insensitive to $J_{-21}$, so changing this parameter only affects high column gas. Temperature is very sensitive to density, however, so that increasing $\Omega_b$ causes large errors.

We have modified the flux in our simulations to fit opacity data better. This could have a negative impact on the accuracy of the results. We have argued previously in section 8. that the effects should be limited to high columns because the equilibrium temperature is very insensitive to the exact flux except at high columns. In figure 4.32 we show this visually. For the upper left panel and the lower right panel $J_{-21} = 0.5$ was used during the simulation. For the lower left panel $J_{-21} = 0.2$ was used during the simulation. similar to the value $J_{-21} = 0.22$ using during the processing to create the figure. Between the upper and lower left panels the only difference is in the black and white contours, which correspond to $N_{HI} = 10^{15} \text{ cm}^{-2}$ and $10^{17} \text{ cm}^{-2}$, respectively. At mean density or greater (responsible for all detectable HI Lyman-$\alpha$ clouds at $z = 3$), the temperature is close to equilibrium and the heating history is relatively unimportant. Cooling times are short and only shock heating takes the gas out of equilibrium. Above overdensities of 10, the equilibrium temperature can vary significantly at $z = 3$. This allows for a lower Jeans mass ($\propto T^{1/2}$) and thus more for small objects to collapse in gas or to collapse to a greater extent by a given time. The lower right panel was evolved with fixed temperatures. The major effect is that the voids are emptier of neutral gas. This is because the gas is much colder in voids normally and with the high
Fig. 4.32.— $N_{HI}$ Column density through 2 Mpc thick slices at $z = 3$. These show the effect of adjusting the temperature history. The initial conditions are the same ($\nu = 0$, SCDM) and the flux level used for the output was $J_{-21} = 0.22$ for all panels except the top right which used $J_{-21} = 0.88$. The upper left panel is the standard simulation; run with $J_{-21} = 0.5$. The upper right panel was run with twice the standard $\Omega_b$ and $J_{-21} = 2.0$. According to this combination and the standard scaling law – the upper two panels should look the same. The lower left panel was run with $J_{-21} = 0.2$. The lower right panel was run with a fixed temperature of 20,000 K. The contours in $\log_{10} N_{HI}$ are 11-12.75 (dark), 13 (dark grey), 13.5 (medium grey), 14 (light grey), 14.5 (dark), 15 (black) and 17 (white).
temperature it is more ionized.

Figure 4.33 shows the effect on the column density distribution. As discussed only high columns are significantly affected. This figure gives us confidence that our flux rescaling procedures have not had any negative effects of our Lyman-α forest results. Simulators taking aim at Lyman limit systems should take note, however.

The isothermal simulation included in the figure shows that the hump in the column density distribution is strongly related to the temperatures associated with different densities. In the presence of uniform temperatures no hump exists. This hump is apparent in the equilibrium temperature versus density plots. This connection has not been made previously: The strong density/column density correlation exhibits itself in the column density distribution via the strong temperature/density relation. The effect is visible as a break in the distribution.

Denser gas generally has higher equilibrium temperatures. The neutral fraction is thus lower for the artificial spectra where $\Omega_b$ is doubled even if $J_{-21}$ is raised to compensate. This is clearly evident in the figures. The amount of neutral gas is reduced everywhere in figure 4.32. This results in a substantial decrease in the flux decrement measured, as seen in section 4.. The line counts are significantly depressed and the hump has moved to the left slightly. The hump in the column density distribution is positioned by the temperature/column density relationship. We can modify the temperature by changing the mean baryon density and thus we can move the hump to different columns. The column density distribution changes differently when ionizing flux rather than baryon density is changed. If the flux is lowered to fit the flux decrement observations the hump does not quite move to its usual position. With the current accuracy of the observation data, we can not exploit this difference. Cosmic and ionizing flux level variations along and between different lines of sight make it difficult to use very fine differences even with all the systematic Voigt fitting differences resolved. The scaling relation, when the temperature variation is taken into account, has a different $\Omega_b$ dependence. Our simulations indicate that $T \propto n_b^{0.65}$. 
Fig. 4.33.— $N_{HI}$ column density distribution at $z = 3$ resulting from variations in the temperature history of the gas. The ionizing flux during the processing was $J_{-21} = 0.19$ and that used during the runs is as marked except that the $J_{-21} = 0.77$ marked for the $\Omega_b = 0.1$ run is the processing value. The curve labelled $T=20000$K was run with uniform fixed temperatures. The isothermal and $\Omega_b = 0.1$ results used $J_{-21} = 0.5$ and $J_{-21} = 4.0$ during the runs respectively.
so a more appropriate scaling for $\tau$ is then:

$$\tau \approx (\Omega_b h^2)^{(2-0.65\alpha_{\text{rec}})}/(J_{-21} H(z)).$$

(4.12)

This scaling is not perfect because the dependences of recombination rate on temperature and equilibrium temperature on density are not strict power-laws. Equation 4.12 is very good near the densities responsible for $N_{HI} \sim 10^{14} \text{cm}^{-2}$ absorbers: $\delta \rho/ \rho > 10$. In figure 4.34 we have plotted the column through the $\Omega_b = 0.1$ simulation against the column through the standard $\Omega_b = 0.05$ simulation using this new relation to set $J_{-21}$ with $\alpha_{\text{rec}} = 0.65$. Differences are observable in the voids and the peaks, but the mean density material is now very similar.

10. Zel’dovich maps

The Zel’dovich mapping described in Chapter 3 can be used to take the initial conditions of a simulation and generate a reasonable facsimile of the numerical result. It shows where the largest structures will tend to be. The mapping is accurate only for small displacements. In a truncated Zel’dovich map the initial conditions are smoothed so that small-scale waves that would otherwise go non-linear at the redshifts of interest are given lower amplitudes. Hui et al. (1997) used truncated Zel’dovich maps combined with the assumption that the gas follows the dark matter, painted on gas temperatures using a fairly good prescription, and then generated artificial spectra from the results. We created truncated Zel’dovich maps of the $\nu = 0$ SCDM initial condition and plotted them in neutral hydrogen against the simulation in figure 4.35. The approximate results do not match the results of full simulations, especially for the large smoothing employed by Hui et al., who were forced to use this value so that dynamically hot regions were appropriately smoothed. The $\nu = 0$ case used for the comparison is not dynamically hot.

11. Helium Absorption
Helium absorption was not a particular focus of this work but will be the subject of future effort. Because significant Helium II absorption can arise at lower densities than those typical of the hydrogen Lyman-α forest, non-equilibrium abundances and the heating history are important. The cooling time for low density gas approaches the Hubble time and thus it has a memory of its previous heating and the temperature is not in equilibrium. This is apparent in figure 4.6 at low densities.

The Helium II Lyman-α opacity is fundamentally dependent on the smaller scale fluctuations than HI. Resolving the Jeans scale does not guarantee that all gas structures are resolved. Dark matter fluctuations can grow independently of the gas and form deep potentials into which the gas can collapse. Higher resolution simulations may be required to draw strong conclusions regarding the source of helium absorption.

Helium column as a ratio of hydrogen column is fairly constant with temperature and density except in two regimes. For our spectral choice the typical $N_{\text{HeII}}/N_{\text{HII}}$ ratio is 30–40. Near dwarf galaxies undergoing mergers, the temperatures jump to a few million degrees. In these regions collisional ionization becomes more important and the ratio rises because hydrogen is more affected. Typical ratios in shocking gas squeezed between merging dwarfs are around 50–80. Within the cool cores of objects, helium recombines sooner than hydrogen and its column levels off at a fixed value, so the ratio becomes very small. For totally neutral gas the ratio drops to the ratio of hydrogen to helium nuclei (1/12). Helium absorption is far too optically thick for columns of individual systems to be measurable. The exact ratios are sensitive to details of the ionizing spectrum which we are unable to observe accurately.

In figure 4.36 we compare the HeII Lyman-α absorption through a 2 Mpc comoving slab with the absorption from HI Lyman-α. This figure was produced for the flux level used during the simulation $J_{-21} = 0.5$ with no modification of the hardness of the spectrum to produce more HeII. The simulated results give a mean optical depth to HeII Lyman-α of $\tau_{\text{HeII}} \sim 0.5$ at $z = 3$ with $J_{-21} = 0.5$. If the flux
is rescaled to match the HI opacity of Rauch et al. (1997b), then \( \tau_{\text{HeII}} = 1.3 \) at \( z = 3 \) (\( J_{-21} = 0.2 \)). Conservatively, HST results indicate values for the optical depth \( \tau_{\text{HeII}} \gtrsim 1.3 \) but the value is likely to be much higher (Hogan et al. (1997)). Hogan et al. and Reimers et al. (1997) estimate \( n_{\text{HeII}}/n_{\text{HI}} \gtrsim 100 \) via comparisons with HI Lyman-\( \alpha \) forest spectra. We find a lower ratio can give results consistent with \( \tau_{\text{HeII}} \sim 1.3 \) but not greater opacity. This implies that the true ionizing spectrum is much softer than the one used. The issue is complicated by the possibility of large-scale spatial variations in the hardness of the spectrum associated with proximity to ionizing radiation sources.

In figure 4.37, in addition to rescaling the overall flux amplitude to match the HI flux decrement, the flux at the helium edge has been softened by a factor of four relative to our initial ionizing spectrum choice. This is definitely more consistent with the observations. This lowers \( J_{24.6\text{eV}} \) from 0.08 to 0.02. The slope of the ionizing spectrum towards higher energies has a small impact (< 20%). Figure 4.38 compares artificial HI Lyman-\( \alpha \) and HeII Lyman-\( \alpha \) spectra before and after the softening. Simulation data from the \( \nu = 0 \), standard cold dark matter run was used. The ratio \( n_{\text{HeII}}/n_{\text{HI}} \sim 120 \) and \( \tau_{\text{HeII}} = 2.3 \).

12. A Cloud/Inter-Cloud split

We find a very strong association between dwarf objects and absorbers with column densities \( N_{\text{HI}} \gtrsim 10^{14.5} \text{ cm}^{-2} \).

Figure 4.39 and figure 4.40 show the structures associated with gas at different overdensities. Bond et al. (1996) predict that overdensity ranges of a few and 5-10 correspond to walls and filaments. The lower two panels of figure 4.40 clearly show that filaments go from fingers pointing toward other peaks to being bridging structures between overdensity 5 and 10.

Figure 4.41 shows the column densities results from parts of the medium with different overdensities. While barely over-dense material (\( \delta < 20 \)) obviously
contributes minimally to high high columns, the contribution from material with \( \delta > 20 \) to these columns seems less than the total. This is an artifact of the fitting procedure because some lines identified as high column are blends. Another factor is the sharp cut we performed in density. Gas that is associated with collapsed objects is excluded. This is expected to be a lesser effect. From these crude investigations it is clear that a split between filaments and cloud objects is promising and consistent with the data. From this study we infer that at \( z = 3 \), the dominant absorbers are dwarf galaxies for neutral hydrogen columns \( N_{HI} \sim 10^{15} - 10^{17} \text{cm}^{-2} \) and filamentary gas below \( N_{HI} \sim 10^{15} \text{cm}^{-2} \). We ran a group finder on the simulation with a \( \delta > 100 \) cut. The smallest objects found typically had velocity dispersions around \( 30 \text{ km s}^{-1} \). There were many fewer objects with larger velocity dispersions (e.g. \( 90 \text{ km s}^{-1} \)).

When we examined the same initial conditions simulationed with half the resolution, only the \( \sim 90 \text{ km s}^{-1} \) objects were resolved consistently. The objects found all have cores giving rise to \( N_{HI} \gtrsim 10^{17} \text{cm}^{-2} \) absorbers. Objects with \( \delta > 20 \) have cores giving rise to \( N_{HI} \gtrsim 10^{14.5} \text{cm}^{-2} \) absorbers.

The onset of 3-dimensional simulations led to a complex picture of Lyman-\( \alpha \) absorbers. The descriptions invoked ideas of filaments and complex webs of structure, overshadowing existing ideas of clouds as discrete objects. While filaments are clearly responsible for the low column absorbers, collapsed objects dominate the moderate to high column absorption.
Fig. 4.34.— Column density through 2 Mpc thick slices at $z = 3$. Left panel: $J_{-21} = 0.22$, $\Omega_b = 0.05$. Right panel: $J_{-21} = 0.63$, $\Omega_b = 0.1$. The contours in $\log_{10} N_{HI}$ are 11-12.75 (dark), 13 (dark grey), 13.5 (medium grey), 14 (light grey), 14.5 (dark), 15 (black) and 17 (white).
Fig. 4.35.— Column density, $N_{HI}$, through 2 Mpc thick slices for the $\sigma_8 = 0.67$ sCDM cosmology are compared with truncated Zel’dovich maps at $z=3$ for various smoothings. The (SPH) smoothing choices are $h/h_{NL}=0.33, 0.125, 0.78$, where $h_{NL} = 1.0$ Mpc. The highly smoothed 0.78 case corresponds to the Hui et al. 1997 choice. These panels show that a locally adaptive smoothing, allowing for smaller values in regions where the dynamical action is not as great, is preferable. The contours in $\log_{10} N_{HI}$ are 11-12.75 (dark), 13 (dark grey), 13.5 (medium grey), 14 (light grey), 14.5 (dark), 15 (black) and 17 (white).
Fig. 4.36.— Column density, $N_{\text{HeII}}$, and $N_{\text{HII}}$, through 2 Mpc thick slices for $z = 3$. The cosmology is standard CDM with $\nu = 0$. The upper left and right panels show $N_{\text{HII}}$ and $N_{\text{HeII}}$ respectively. The contours in $\log_{10} N_{\text{HI}}$ are 11-12.75 (dark), 13 (dark grey), 13.5 (medium grey), 14 (light grey), 14.5 (dark), 15 (black) and 17 (white). The lower panels show the ratio $N_{\text{HeII}}/N_{\text{HII}}$. The lower right panel zooms in on the lower left. For the lower panels the shades mark ratios starting at 0 (black) 10 (dark grey) 30 (medium grey - dominant) 40 (lighter grey) 50 (lightest grey) and 70 (white).
Fig. 4.37.— Column density, $N_{\text{HeII}}$ and $N_{\text{HI}}$, through 2 Mpc thick slices for $z = 3$. The cosmology is standard CDM with $\nu = 0$. The flux has been rescaled to fit the HI opacity and lowered at the HeII edge. The upper left and right panels show $N_{\text{HI}}$ and $N_{\text{HeII}}$ respectively. The contours in $\log_{10}N_{\text{HI}}$ are 11-12.75 (dark), 13 (dark grey), 13.5 (medium grey), 14 (light grey), 14.5 (dark), 15 (black) and 17 (white). The lower panels show the ratio $N_{\text{HeII}}/N_{\text{HI}}$. The lower right panel is a zoom in of the lower left. For the lower panels the shades mark ratios starting at 0 (black) 40 (dark grey) 120 (medium grey – dominant) 160 (lighter grey) 200 (lightest grey) and 280 (white). Note the factor of 4 over the previous figure.
Fig. 4.38.— Sample HeII Lyman-α spectra overlaid with HI Lyman-α spectra arising in the same gas. The top set of eight panels was generated with fluxes satisfying the HI opacity observations only. In the lower panels the spectrum has been softened to better match the HeII observations.
Fig. 4.39.— Column density, $N_{HI}$, through 2 Mpc thick slices for four cosmologies with UV flux scaled to fit the observed flux decrement one-point distribution function. $\Omega_b h^2=0.0125$ throughout. Top left: SCDM $\nu = 0$ run, $z=3$. Top right: with only $\delta < 20$ gas. Bottom left: Dark matter to trace out the halos (log dark matter density contours). Bottom right: with only $\delta > 20$ gas. The contours in $\log_{10} N_{HI}$ are 11-12.75 (dark), 13 (dark grey), 13.5 (medium grey), 14 (light grey), 14.5 (dark), 15 (black) and 17 (white).
Fig. 4.40.— The Lyman-α intercloud medium: Column density $N_{HI}$, through 2 Mpc thick slices for four cosmologies with UV flux scaled to fit the observed flux decrement one-point distribution function. $\Omega_{m} h^2 = 0.0125$ throughout. Top left: SCDM $\nu = 0$ run, $z=3$ with only the $\delta > 80$ cloud contribution shown. Top right: at $\delta > 2$, the intrafilament webbing appears. The web theory of filaments predicts that the typical smoothed overdensity should be $\sim 5 - 10$. Bottom right: $\delta > 10$ gas. Bottom left: $\delta > 5$ gas. A similar story holds for other cosmologies. The contours in $\log_{10} N_{HI}$ are 11-12.75 (dark), 13 (dark grey), 13.5 (medium grey), 14 (light grey), 14.5 (dark), 15 (black) and 17 (white).
Fig. 4.41.— cloud/intercloud split in columns. The Column density distribution per unit redshift at $z = 3$. The observational data are plotted as boxes with widths equal to the bin widths and top and bottom edges at the positions of the 1 $\sigma$ poisson error bars.
REFERENCES


12th Kingston Meeting on Theoretical Astrophysics. in preparation. (BW)


Eds. P. Petitjean and S. Charlot


Steinmetz, M. 1996, to be published in The Early Universe with the VLT, ESO Workshop


Conclusion

This thesis set out to create a set of simulations of the Lyman-\(\alpha\) forest for multiple cosmologies that simultaneously achieved superb high resolution and included the necessary large scale power. We have achieved this with simulations whose results fit the quasar absorption observations in close detail: the flux decrement distribution (figure 4.11), the column density distribution and its evolution (figure 4.22) and the distribution of line widths (figure 4.25). Our fundamental result is that cosmological models with similar fluctuation amplitudes at small scales fit all aspects of the absorption data for \(N_{HI} \lesssim 10^{16} \text{cm}^{-2}\) lines. Specifically, the open, vacuum, hot plus cold and standard cold dark matter models. that are consistent with intermediate scale power-spectrum constraints. reproduce the Lyman-\(\alpha\) forest observations successfully, with values of the baryon density and ionizing background flux well inside observational and theoretical limits.

We show that the correct scaling for the optical depth to Lyman-\(\alpha\) absorption is \(\tau \propto \Omega_b^{1.5} h^3/(J_{-21} H(z))\) in terms of the baryon density \(\Omega_b\) and the ionizing background radiation level \(J_{-21}\). By modifying \(\Omega_b\) or \(J_{-21}\). we can exactly fit each cosmology to the observed flux decrement distribution \(P(1 - e^{-\tau})\). The optical depth does not scale as \(\Omega_b^2\), because the temperature is insensitive to the ionizing flux level but sensitive to density changes. The temperature–density relation creates the evolving break in the column density distribution, observed by Kim et al. (1997).

For \(\Omega_b h^2 = 0.0125\), all cosmologies fit the flux decrement observations with \(J_{-21} \sim 0.2\) at \(z = 3\). Scaling to \(\Omega_b h^2 = 0.02\) to match an estimate by Fukugita et al. (1998), we get \(J_{-21} \sim 0.4\), close to the proximity effect estimate of Giallongo et al. (1997). This is remarkably good agreement.

All the simulated results depend sensitively on the highly uncertain ultraviolet background. We demonstrate that data such as the observed flux decrement must be used to set the ultraviolet levels for each cosmology individually before meaningful comparisons may be made for other data. Consequently, we show that the column
density distribution is an insensitive cosmological probe. Even drastically changed power spectrum amplitudes successfully fit the column densities, though they fail to fit the one-point flux decrement distribution.

Our simulated helium results fit the data convincingly with a soft ionizing flux:
\[ \frac{J_{24.6\text{eV}}}{J_{13.6\text{eV}}} \lesssim 0.1 \text{ at } z = 3. \] This is very soft, closer to a stellar spectrum than to the standard quasar power law.

We show that low resolution Lyman-\( \alpha \) simulations fail to resolve the smaller dwarf galaxies and filaments that should dominate the intermediate and low column absorption respectively. Consequently, they under-predict the flux decrement and the line counts. The pressure of the gaseous medium is shown to set a minimum scale that we resolve.

There are strong visual differences between the cosmologies related to the relative amounts of large scale power (figure 4.13). To differentiate we must seek measures that probe large scales: clustering of clouds (\textit{e.g.} Meiksin and Bouchet 1995), cloud pairing in multiple lines of sight (\textit{e.g.} Fang \textit{et al.} 1996) and other two-point statistics. As the constraints improve, differences between viable models will become increasingly subtle. In future work we will make use of available high quality data (\textit{e.g.} Kirkman and Tytler 1997) to extract high order pixel-pixel statistics for objective comparisons between simulations and observations.

Our investigations show a clear cloud/intercloud split: at \( z = 3 \), the dominant absorbers are dwarf galaxies for neutral hydrogen columns \( N_{HI} \sim 10^{15} - 10^{17}\text{cm}^{-2} \) and filamentary gas below \( N_{HI} \sim 10^{15}\text{cm}^{-2} \). We plan to use group finding and focused simulations to investigate this split further. The split confirms the insight behind the mini-halo models of Rees (1986) and Bond \textit{et al.} (1988) and explains why columns above \( N_{HI} \sim 10^{15}\text{cm}^{-2} \) always have associated metal lines: the dwarf galaxies have dense gaseous cores (producing Lyman limit absorption) and they are forming stars. Accurate future models must include star formation (\textit{e.g.} Klypin 1998) and radiative transfer (\textit{e.g.} CLOUDY, Ferland 1989).
We envisage a strong relationship between the mass or velocity dispersion of these collapsed objects and their cross-sections for Lyman-α absorption at different columns. If future work we hope to establish such an invaluable relationship so that we can tie peak patch objects to absorbing systems. Our current sets of simulations provide a complete database of collapsed objects whose properties could be measured and directly linked to those measured by peak patch.

The peak patch method rapidly produces catalogues of collapsed objects. We have described how we can tile space with multiple peak-patch realisations to probe structure on scales beyond the dynamic range of full numerical simulations. We presented an example of this for galaxy catalogues (modelling the Steidel (1997) cluster at $z = 2.9$, figure 3.8). We are currently working to apply peak patches to reproduce the spatial distribution of Lyman absorbers. There is a lot of observational data accumulating without much theoretical response. We now have a tool built for the problem.

We advocate the use of constrained initial conditions in theoretical studies. We have clearly demonstrated their benefits for Lyman-α forest simulations: with modest computing resources we have uncovered important results while other groups employ orders of magnitude more computing power to achieve results that are still inferior in regard to the inclusion of large scale power (figure 3.2).

In other work, we are estimating the observability in X-rays, the Sunyaev-Zel’dovich effect and gravitational lensing of the large scale cosmic web with constrained super-cluster patches. Using a magneto-hydrodynamics code written with J. Murray, we simulated the growth of cosmological magnetic fields in galaxy-galaxy filaments and are about to investigate the role of magnetized filaments in molecular cloud complexes. These projects benefit from the precision of constrained initial conditions.

The future for constrained, non-periodic simulations is bright. We can focus our computing power exactly where we want it: high resolution. The initial condition
constraints are posed in terms of variables with well-understood probability distributions, allowing us to place the results into the cosmological context. Combining tiled peak patch realisations with suites of high resolution studies tailored to the patches, we have a tool to study the large-scale spatial distribution of Lyman-α absorbers, galaxies or any other important astrophysical population.