Title
The Effect of Inhomogeneous Helium II Reionization on the Lyman Alpha Forest

Permalink
https://escholarship.org/uc/item/79b2c99g

Author
Chen, Pengfei

Publication Date
2016

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

The Effect of Inhomogeneous Helium II Reionization on the Lyman Alpha Forest

A Dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Physics (with a specialization in Computational Science)

by

Pengfei Chen

Committee in charge:

Professor Michael Norman, Chair
Professor Gert Cauwenberghs
Professor Philip E. Gill
Professor Dusan Keres
Professor David Tytler

2016

The Dissertation of Pengfei Chen is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

Chair

University of California, San Diego

2016

DEDICATION

To

My Grandparents

For being my first teachers

My Parents

For their unconditional love

Jie

For her love and support

EPIGRAPH

He has made everything beautiful in its time.

He has also set eternity in the human heart;

yet no one can fathom what God has done from beginning to end.

Ecclesiastes 3:11

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

Special thanks to my advisor, Michael Norman. For the past 5 years, he has been so helpful and patient. I'm very grateful for his guidance, encouragement, and support. Many thanks to David Tytler for his support and enlightening discussions.

I would like to thank all my committee members, Gert Cauwenberghs, Philip E. Gill, Dusan Keres, David Tytler for their time and comments.

Many thanks to Hao Xu and Geoffrey So, for helping me out so many times, and for so many useful and interesting discussions.

Many thanks to John Wise for helping me so many times. Many thanks to Daniel Reynolds for his MGFLD radiation transfer algorithm, without which I could not complete this work. Many thanks also go to Mahidhar Tatineni, for his technical support on Comet.

I would also like to thank the community of yt, Enzo, and Blue Waters. They have been so helpful and professional. Thanks also go to all professors who have taught me and others who helped me during my graduate study.

Last but not least, many thanks to my family members and Jie, for their support, understanding, and love.

Chapter 7, in full, is currently being prepared for submission for publication with the title "Fully Coupled Simulation of Cosmic Reionization. III. Contribution of the Smallest Galaxies". Authors are Chen, Pengfei; Norman, Michael; Xu, Hao; Wise, John. The dissertation author was the primary investigator and author of this material.

2010	Bachelor of Science in Physics, University of Science and Technology of China
2010-2011	Graduate Student, Indiana University, Bloomington
2014	Master of Science in Physics, University of California, San Diego
2016	Ph. D. in Physics with specialization in Computational Science, University of California, San Diego

PUBLICATIONS

Pengfei Chen, John Wise, Michael Norman, Hao Xu, Brian O'Shea, "Scaling Relations for Galaxies Prior to Reionization", *The Astrophysical Journal*, 795, 144, 2014.

ABSTRACT OF THE DISSERTATION

The Effect of Inhomogeneous Helium II Reionization on the Lyman Alpha Forest

by

Pengfei Chen

Doctor of Philosophy in Physics (with a specialization in Computational Science)

University of California, San Diego, 2016

Professor Michael Norman, Chair

The Lyman Alpha Forest (LAF) has been extensively studied observationally and by simulations. Good agreement between data and simulations has been reached in some statistics describing the LAF. However, there are still significant discrepancies in the magnitude of the flux power spectrum (FPS) and the shape of the flux probability distribution function (FPDF). Here we explore a possible resolution by adding new physical ingredients into the standard LAF simulations.

Firstly, we repeat the standard simulation. Here the intergalactic medium (IGM) is assumed to be photoionized and photoheated by a homogeneous ultraviolet

background. Simulations with various box sizes and He II photoheating rates all confirm the mismatch.

Then we introduce simulations that treat quasars as point sources, and calculate the radiation transfer using multigroup flux-limited diffusion. Our simulations here fall into two categories. Simulations with larger volumes $(126 \text{ Mpc}/h)^3$ sample the more luminous quasars and exhibit a more reliable thermal and ionization history. Due to the finite time it takes He III bubbles to expand, our simulations reionize later $(z \sim 2.76)$ than standard simulations, and show better agreement with observations.

Simulations with higher spatial resolutions $(80 \text{ Mpc}/h)^3$ are used to generate synthetic spectra. Our mean H I effective optical depths and rescaled b_σ's agree with data well, but the discrepancies in FPS and FPDF still exist. We therefore conclude that the inhomogeneous heating due to He II reionization is not the missing physics needed to reconcile simulations with data. We also generate He II LAF spectra and successfully reproduce the large scatter in the He II effective optical depth before He II reionization completes. We also calculate its FPS and FPDF, which could be used to compare with future observations.

On a related topic, we conduct simulations of the hydrogen reionization to examine the effect of the smallest galaxies. We find that reionization begins substantially earlier when galaxies in the mass range $10^7 \leq M_{dyn}/M_\odot \leq 10^8$ are included, but that the redshift of reionization completion is little affected. This approach could be combined with He II reionization simulations discussed above to provide a more complete model of the cosmic thermal history.

Chapter 1

Introduction

1.1 Lyman Alpha Forest

Quasars, or quasi-stellar radio sources, are extremely luminous objects with very broad emission lines (Greenstein and Schmidt, 1964). Because they are so bright, they are visible to redshifts as high as z=7, making them excellent cosmological probes (Fan et al., 2005). The Sloan Digital Sky Survey (SDSS) has catalogued hundreds of thousands of quasars (http://www.sdss.org/dr12/algorithms/boss-dr12-quasar-catalog/). It's now believed that each of them is a compact region surrounding a supermassive black hole in the center of a massive galaxy (Kauffmann and Haehnelt, 2000). The Lyman Alpha (Lyα) Forest (LAF) is a group of absorption lines observed in the spectrum of distant quasars (e.g. Fig. 1.1). The absorption happens when the electron of neutral hydrogen in the intergalactic medium (IGM) goes from n=1 to n=2, where n is the principle quantum number. The rest wavelength of Lyα line is 1216Å, which is in the ultraviolet range of electromagnetic spectrum. As the emitted light from quasars intersects with multiple clouds of neutral hydrogen
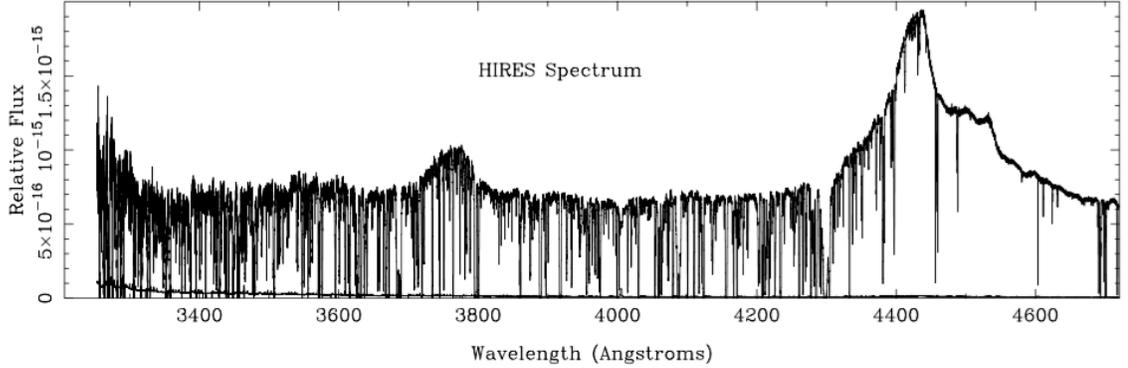
Figure 1.1: Spectrum of quasar HS 0105+1619 at redshift $z = 2.64$ from O'Meara et al. (2001). It's a flux-calibrated HIRES spectrum. The rest wavelength of Lyα line is 1216Å, but the observed wavelength of HS 0105+1619 is 1216Å×(1+2.64)=4426Å due to redshift. The LAF here is the absorption lines to the left of 4426Å, which correspond to clouds between the quasar and us.

in the expanding universe, each cloud leaves a fingerprint as an absorption line in the spectrum we observe at wavelength 1216Å(1+z), where z is the redshift of the absorbing cloud (Sargent et al., 1980). In the mid 1990s hydrodynamic cosmological simulations revealed the physical origin of the LAF (Rauch, 1998). Namely, each LAF absorption feature corresponds to the quasar sightline intersecting the "cosmic web" of baryons tracing the dark matter filaments which are the result of structure formation. The gas is in a highly ionized state due to the strong ultraviolet (UV) background from galaxies and quasars (Haardt and Madau, 2012). The LAF is an important tool to determine the distribution of neutral hydrogen clouds, the ionization history of universe, and to constrain cosmological parameters (Weinberg et al., 2003).

An important observable of the LAF is effective optical depth τ^{eff}, which is defined as $e^{-\tau^{\mathrm{eff}}} = \left\langle \frac{F_{\mathrm{obs}}}{F_{\mathrm{cont}}} \right\rangle$, where F_{obs} is the observed flux, and F_{cont} is the estimated unabsorbed flux (Oke and Korycansky, 1982). τ^{eff} could be measured with low resolution spectroscopy (Rauch, 1998). Fig. 1.2 shows measurement results for H I

and He II Lyα τ^{eff} as a function of redshift.

Keck HIRES and VLT UVES spectroscopic observations have provided high-resolution spectra (FWHM < 25 km/s) on the LAF from $z \sim 2$ to $z \sim 4$. The Lyα line shapes are found to be well approximated by Voigt profiles (e.g. Carswell et al. (1984)). By fitting high-resolution spectra with a Voigt profile, one can obtain the line width, i.e. Doppler parameter b. It can be decomposed into: $b = \sqrt{b_{\mathrm{th}}^2 + b_{\mathrm{turb}}^2 + b_{\mathrm{Hubb}}^2}$, where $b_{\mathrm{th}} = \sqrt{\frac{2kT}{m}}$ is the thermal broadening of the absorption line, b_{turb} is the turbulent contribution to the line width (Rauch, 1998), b_{Hubb} is the contribution due to Hubble broadening across the width of the absorbing filament (Bryan et al., 1999). So b could give us information about the temperature and dynamics of Lyα absorbers. Other observables and discrepancies between observations and simulations are described in Section 1.4.

Hydrodynamical cosmological simulations have been conducted to study the properties of the IGM and LAF. Synthetic LAF spectra could be produced by shooting light rays in the simulation box and taking neutral hydrogen density, temperature, Hubble constant and peculiar velocity along line of sight into calculation (Cen et al., 1994; Zhang et al., 1997). More recent studies have examined the dependence of the LAF observables on cosmological parameters (Jena et al., 2005), box size (Tytler et al., 2009), and numerical resolution (Bryan et al., 1999).

1.2   He II reionization

The epoch of reionization is the period of time when the IGM transformed from a primarily neutral state to a nearly fully ionized state. The ionization potentials of H I, He I and He II are 13.6 eV, 24.6 eV and 54.4 eV, respectively. Given those
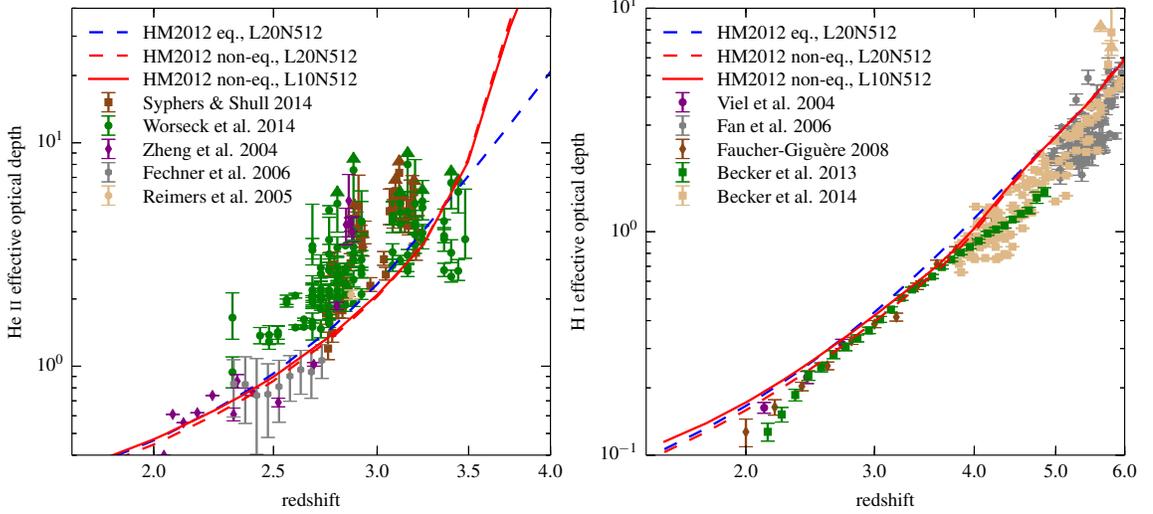
Figure 1.2: Left: Effective optical depths for He II Lyα absorption versus redshift from Puchwein et al. (2014). The results (lines) are based on equilibrium and non-equilibrium simulations with the UV background from Haardt and Madau (2012). Observational data from Zheng et al. (2004), Reimers et al. (2005), Fechner et al. (2006), Syphers and Shull (2014) and Worseck et al. (2014) are shown for reference. Right: Effective optical depths for H I Lyα absorption as a function of redshift from Puchwein et al. (2014). Observational data from Viel et al. (2004), Fan et al. (2005), Faucher-Giguere et al. (2008), Becker et al. (2012) and Becker et al. (2014) are shown. The x-axis is linear in $\log(1 + z)$ so that a power-law evolution corresponds to a straight line.

different ionization potentials, the reionization of different species happened in different period of time depending on the population of ionizing sources with different spectral hardness. In the standard picture of reionization, H I and He I in IGM was reionized by UV light emitted by early star-forming galaxies by $z \sim 6$, as evidenced by the transmission of H I Lyα photons in quasar spectra up to $z \sim 6$ (e.g. Becker et al. (2001); Fan et al. (2005)). But those sources are not efficient in ionizing He II due to the higher ionization potential.

It's believed that He II was ionized later by quasars, when they were abundant enough to provide hard photons (Madau and Meiksin, 1994; Miralda-Escudé et al., 2000). Observations of high-redshift quasars show a substantial increase in the mean He II effective optical depth $\left\langle \tau_{\mathrm{HeII}}^{\mathrm{eff}} \right\rangle$ from $z \sim 2.3$ to $z \sim 3.2$, but with a large scatter at $2.7 < z < 3$ (see Fig. 1.2 Left). Troughs with large optical depth ($\tau_{\mathrm{HeII}}^{\mathrm{eff}} > 4$) on scales of ~ 10 proper Mpc and windows of flux transmission have been observed in multiple sightlines at $2.7 < z < 3$, which indicates that He II reionization was an extended process and was not complete until $z < 2.7$ (Shull et al., 2010; Worseck et al., 2011).

1.3 How He II reionization affects the LAF

The heat input of the photoionization of He II may affect the thermal balance of the IGM, which would affect the IGM temperature and hence the statistics of the H I LAF. McQuinn et al. (2009) ran a set of large scale simulations of He II reionization and found that He II reionization heats IGM regions by as much as $25,000$ K, with the volume-averaged temperature increasing by $\sim 12,000$ K.

However, getting reliable measurements of the IGM temperature has been a challenging task. Schaye et al. (2000) measured the line widths of the H I LAF

observations and found that the temperature at the mean density has a peak at redshift $z \sim 3$ and the gas of different densities becomes nearly isothermal then. Theuns et al. (2002) found a dip in $\left\langle \tau_{\mathrm{HI}}^{\mathrm{eff}} \right\rangle$ at redshift $z \sim 3.1$, and interpreted this as the result of a temperature increase from He II reionization. However, McDonald et al. (2001) and Zaldarriaga et al. (2001) did not find the sudden increase in IGM temperature. Later observations of $\left\langle \tau_{\mathrm{HI}}^{\mathrm{eff}} \right\rangle$ also did not confirm the dip (Kim et al., 2007; Dall'Aglio et al., 2009).

Becker et al. (2011) used a statistic based on the curvature of the H I LAF and converted it into the temperature at the mean density T_0 after calibrating with multiple hydrodynamic simulations. They applied this method to 61 high-resolution quasar spectra and found a gradual increase in T_0 at $2.8 < z < 4.4$, regardless of assumptions for γ in the temperature-density relation (See Fig. 1.3 and Eqn. 3.1). They interpret the rise of the temperature as an evidence of extended He II reionization, which is consistent with the opacity measurements of the He II LAF (See Fig. 1.2 Left).

The small-scale structure of the LAF depends on the thermal broadening, the Hubble broadening and any turbulent broadening on the absorbers. Extra heat from He II reionization may increase the thermal broadening. Higher temperature corresponds to a larger Jeans length which will increase the characteristic size of absorbers. Therefore He II reionization also increases the Hubble broadening of absorbers on average. Furthermore, the increased pressure smooths out small-scale density fluctuations, thus reduces the clumping factor, which makes it easier to keep the IGM ionized (e.g., Pawlik et al. (2009)).
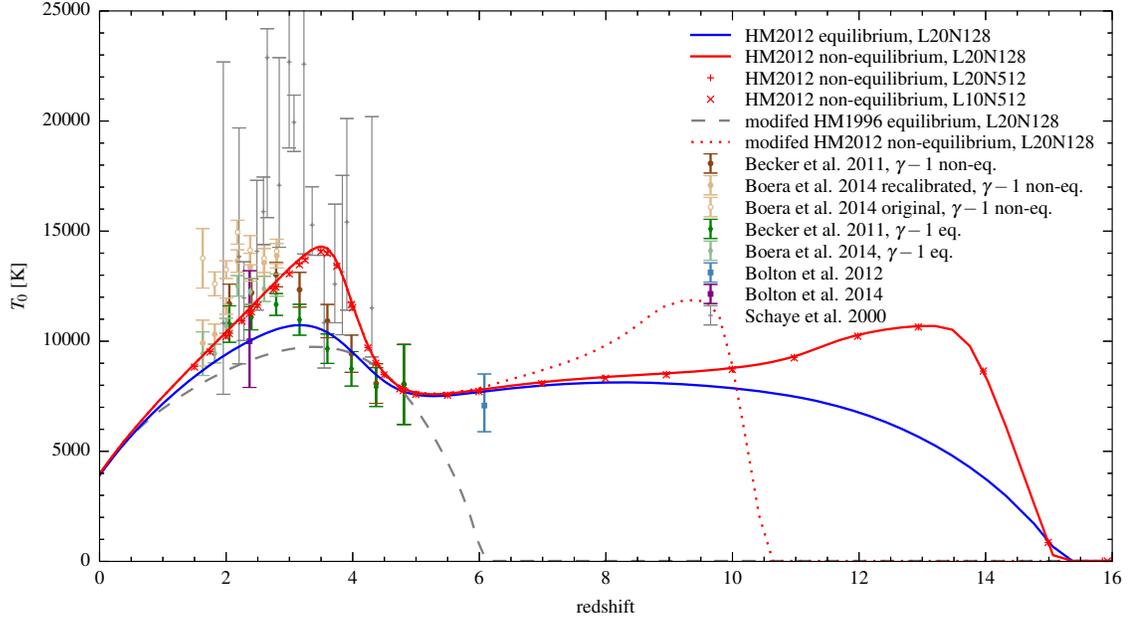
Figure 1.3: The median IGM temperature at mean density $(T_{\mathrm{median}}(\Delta = 1))$ as a function of redshift from Puchwein et al. (2014). Results for equilibrium and non-equilibrium simulations with the HM2012 UV background (Haardt and Madau, 2012) are shown. They also include temperatures in an equilibrium run with a modified background from Haardt and Madau (1995), as well as temperatures in a non-equilibrium run with a modified HM2012 background (see their Section 3.3.1 for details). Observational constraints from Schaye et al. (2000), Becker et al. (2011), Bolton et al. (2012) (excludes He II heating), Bolton et al. (2013) and Boera et al. (2014) are shown for comparison.

1.4 Mismatch between standard simulations and observations

Standard hydrodynamic simulations (Cen et al., 1994; Zhang et al., 1997) have made synthetic LAF and reproduced some of the statistics describing the LAF after adjusting the ionization rate and/or the heat input. Day (2013) conducted a detailed search of the entire parameter space and found that none of those standard simulations can reproduce all key parameters of the LAF simultaneously. The mismatch is in three ways: (1) the power spectrum of the flux in the H I LAF has a wrong shape; (2) the amplitude of the 1D power spectrum is lower than observations on Mpc scales; (3) the distribution of the amount of Lyα absorption in spectra, i.e. the flux probability distribution function (PDF), has a wrong shape (Becker et al., 2007; Tytler et al., 2009).

Tytler et al. (2009) conducted multiple simulations with Enzo but were unable to find a simulation that could simultaneously fit the line widths, the mean absorption by Lyα and the power spectrum of the flux in the LAF. They could fit two of the three, by adjusting the astrophysical parameters, but could not fit all three simultaneously. Fig. 1.4 shows results from one of their simulations that fits both the mean absorption and the line widths in data at $z = 2$. In the left panel the observed power is 50% larger than simulations on large scales ($\log(k/(\text{s/km})) \sim -2$). The right panel shows that simulations also do not reproduce the H I LAF flux PDF. Specifically, their simulations got fewer pixels with a lot of absorption (e.g. flux = 0.15) and more pixels with less absorption (e.g. flux = 0.8) than data from Kim et al. (2007).
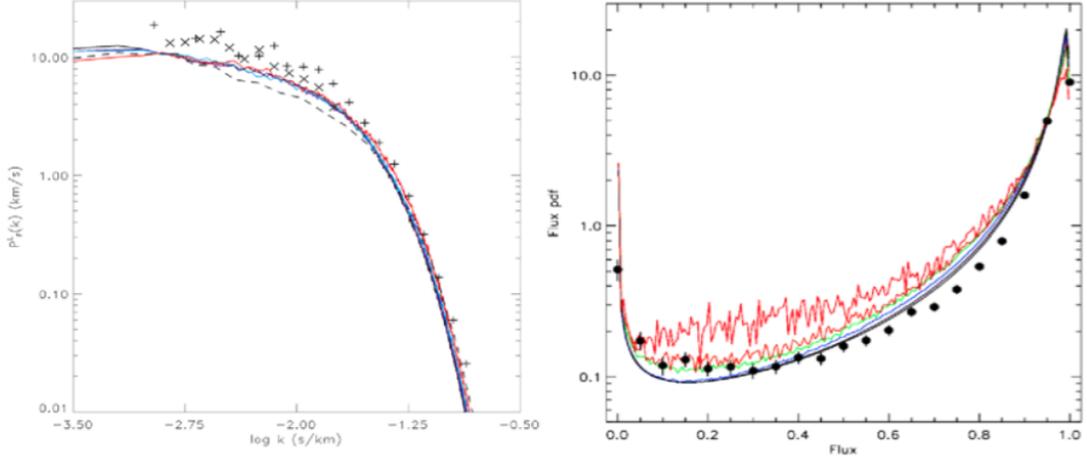
Figure 1.4: Left: Power spectrum of the flux from data (points) and simulations in Tytler et al. (2009) (curves) versus the wave number from Tytler et al. (2009). From left to right, scales decrease from tens of Mpc to sub-Mpc. The power from McDonald et al. (2005) (SDSS results linearly extrapolated to $z = 2$) and Jena et al. (2005) is shown. Increasing the box size has little effect: 76.8 Mpc box (black), 38.4 Mpc (violet), 19.2 Mpc (blue) and 9.6 Mpc (red) solid lines. Right: H I LAF flux PDF from data in Kim et al. (2007) (points) and simulations from Tytler et al. (2009) (curves). Curves are identical except doubling in box size from red (18 Mpc) to orange, green, blue, black (76 Mpc). Simulations have significantly less pixels with low flux values and more pixels with high flux values.

1.5   Simulating He II reionization: challenges

There are several challenges in the simulation of He II reionization. The first one is the large dynamic range it requires. On one hand, quasars have typical separations of tens of comoving Mpc, on the other hand, in order to reproduce the small scale features of LAF, we need to resolve down to tens of comoving kpc. Lukić et al. (2015) conducted a set of simulations covering the redshift range $2 \leq z \leq 4$ and found that a grid resolution of 20 h^{-1} kpc is required to produce 1 per cent convergence of LAF flux statistics, up to scales of 10 h^{-1} Mpc, where $h \approx 0.7$ is Hubble constant in units of 100 km/s/Mpc. They also found that box sizes of 40 h^{-1} Mpc are needed to suppress numerical errors due to missing modes of matter power to a sub-per cent level. Thus a large dynamic range is required to both sample enough quasars and spatially resolve the LAF. Taking the ratio of the inner and outer scales, we see that we need a uniform mesh of approximately 2000^3 grids. Due to the current limited computational power, previous simulations investigating the effect of He II reionization on LAF mainly fall in two categories. Some of them focused on small scales of tens of Mpc to resolve the LAF (e.g. Bolton et al. (2009), Meiksin and Tittley (2012)), but they failed to capture the patchiness of He II reionization process. Others focused on large scales of hundreds of Mpc to sample more quasars (e.g. Paschos et al. (2007), McQuinn et al. (2009)), but they didn't have enough spatial resolution to resolve the details of LAF.

The second challenge in the simulation of He II reionization lies in the radiation transfer. Since UV photons with different energy have different interaction cross section with H I/He I/He II and thus have different mean free paths, it is essential to transport UV photons with different energy separately according to the spectral shape of quasars. This would require multi-group radiation transfer, which would consume

more memory and computation time as the number of energy groups increases.

Another challenge is about the uncertainty of quasar luminosity function (LF). Although it is well measured at low redshift (e.g. Hopkins et al. (2006)), it's poorly constrained at high redshift ($z > 4$) due to the limited number of faint quasars observed at high redshift (Shankar and Mathur, 2007). Giallongo et al. (2015) found 22 active galactic nucleus (AGN) candidates at $z > 4$ and derived a preliminary estimate of the UV LF in the redshift interval $4 \leq z \leq 6.5$. More samples need to be selected to produce a more reliable estimate.

1.6   Plan for this thesis

In this thesis, we use fully coupled simulations including hydrodynamical, gravitational, chemical processes and radiation transfer to study the impact of He II reionization on both the H I and He II LAFs. In Chapter 2 we will describe the methodology used in all subsequent simulations and show some test results. We will then show results of standard optically thin LAF simulations in Chapter 3, i.e. simulations with UV background as the only ionizing source. We then add quasars into our simulations. The methodology of placing quasars and the global evolution results are presented in Chapter 4. The spectral analysis of H I and He II LAFs are discussed in Chapter 5 and Chapter 6, respectively. In Chapter 7 I present research not directly related to the main topic of this thesis; research done somewhat earlier in my graduate career, where I show simulations about hydrogen reionization with both UV background and galaxies as ionizing sources. In Chapter 8 we reach our conclusions and discuss about open questions. We show some of the core codes we use to setup and analyze our simulations in the appendix A.

Chapter 2

Methodology

2.1 Cosmological framework

All simulations in subsequent chapters are cosmological simulations, where we solve the coupled equations of gas dynamics, dark matter dynamics, self-gravity, gas chemistry, radiation transfer and gas cooling/heating in a comoving volume of the expanding universe. Here we only do unigrid simulations, which means all equations we solve are discretized on a uniform Cartesian mesh.

The spacetime model we use is a Friedmann-Robertson-Walker (FRW) model with Lambda Cold Dark Matter (ΛCDM) cosmology. We consider the commonly used "6-species" model of primordial gas (Abel et al., 1997; Anninos et al., 1997). In most of the work here we don't consider star formation, but use ionization background and/or point sources representing first galaxies or quasars. To better compare our results with different models, we choose the same random seed to generate initial conditions every time, so we have the same initial distribution of gas and dark matter. Then, after choosing appropriate boundary conditions, we solve all equations including radiation

transfer self-consistently on the same mesh, instead of post-processing the simulation results to take radiation transfer into account.

2.2 Equations solved

The coupled equations we solve in all our simulations are listed below (Reynolds et al., 2009).

$$\nabla^2 \phi = \frac{4\pi g}{a}(\rho_b + \rho_{dm} - \langle \rho \rangle), \tag{2.1}$$

$$\partial_t \rho_b + \frac{1}{a}\mathbf{v}_b \cdot \nabla \rho_b = -\frac{1}{a}\rho_b \nabla \cdot \mathbf{v}_b - \dot{\rho}_{SF}, \tag{2.2}$$

$$\partial_t \mathbf{v}_b + \frac{1}{a}\left(\mathbf{v}_b \cdot \nabla\right)\mathbf{v}_b = -\frac{\dot{a}}{a}\mathbf{v}_b - \frac{1}{a\rho_b}\nabla p - \frac{1}{a}\nabla \phi, \tag{2.3}$$

$$\partial_t e + \frac{1}{a}\mathbf{v}_b \cdot \nabla e = -\frac{2\dot{a}}{a}e - \frac{1}{a\rho_b}\nabla \cdot (p\mathbf{v}_b) - \frac{1}{a}\mathbf{v}_b \cdot \nabla \phi + G - \Lambda + \dot{e}_{SF} \tag{2.4}$$

$$\partial_t \mathbf{n}_i + \frac{1}{a}\nabla \cdot (\mathbf{n}_i \mathbf{v}_b) = \alpha_{i,j}\mathbf{n}_e\mathbf{n}_j - \mathbf{n}_i\Gamma_i^{ph}, \qquad i = 1, \ldots, N_s. \tag{2.5}$$

Eqn. (2.1) is used to solve the modified gravitational potential ϕ, where g is the gravitational constant, ρ_b is the comoving baryonic density, ρ_{dm} is the dark matter density, and $\langle \rho \rangle$ is the cosmic mean density. We evolve the dark matter density ρ_{dm} using the Particle-Mesh method (Hockney and Eastwood, 1988; Norman and Bryan, 1999; Bryan et al., 2014a). Eqn. (2.2) to (2.4) show the conservation of mass, momentum and energy, respectively (Bryan et al., 1995). $\mathbf{v}_b$ is the proper peculiar baryonic velocity, and it relates the comoving peculiar baryonic velocity $\dot{\mathbf{x}}$ with $\mathbf{v}_b = a(t)\dot{\mathbf{x}}$, where $a(t)$ is the cosmological expansion factor, and it relates redshift z with $a(t) = (1 + z)^{-1}$. $a(t)$ is obtained by solving the Friedmann equation. p is the proper pressure, and e is the total gas energy per unit mass. G is the photoheating

rate and it's defined later in Eqn. (2.11) and (2.21). Λ is the gas cooling rate due to chemical processes and it depends on both the chemical number densities and current gas temperature (Abel et al., 1997; Anninos et al., 1997),

$$T = \frac{2\,p\,\mu\,m_p}{3\,\rho_b\,k_b},\qquad\qquad(2.6)$$

where m_p is the mass of a proton, μ is the local molecular weight, and k_b is Boltzmann's constant. Eqn. (2.5) shows ionization processes between 6 chemical species H I, H II, He I, He II, He III and the electron e^-. $\mathbf{n}_i$ is the ith species's comoving number density, $\mathbf{n}_e$ is the electron number density, $\alpha_{i,j}$ are the reaction rate coefficients for the interactions between species i and j and they are highly temperature dependent (Abel et al., 1997; Hui and Gnedin, 1997), Γ_i^{ph} is the photoionization rate and it's defined later in Eqn. (2.12) and (2.20) (Abel et al., 1997; Hui and Gnedin, 1997). We also use a standard ideal gas equation,

$$e = \frac{p}{2\rho_b/3} + \frac{1}{2}|\mathbf{v}_b|^2.\qquad\qquad(2.7)$$

Additionally, most simulations we conduct solve equations for radiation transfer. Since they differ in different simulations, we discuss them in subsections.

2.2.1 Grey FLD

For simulations in Chapter 3 and Chapter 7, we solve the following equation for the grey flux-limited diffusion (FLD) approximation of radiation transport (Hayes

and Norman, 2003; Paschos, 2005),

$$\partial_t E + \frac{1}{a} \nabla \cdot (E \mathbf{v}_b) = \nabla \cdot (D \nabla E) - \frac{\dot{a}}{a} E - c \kappa E + \eta, \qquad (2.8)$$

where E is the comoving grey radiation energy density, D is the flux limiter defined by (Morel, 2000),

$$D = \min \left\{ c \left(9\kappa^2 + R^2 \right)^{-1/2}, D_{max} \right\}, \quad \text{and} \quad R = \max \left\{ \frac{|\partial_x E|}{E}, R_{min} \right\}. \qquad (2.9)$$

κ is the opacity. $D_{max} = 0.006 \, c \, L_{unit}$, where L_{unit} is the length unit for the simulation. $R_{min} = 10^{-20}/L_{unit}$.

To define the grey radiation energy density $E(\mathbf{x}, t)$, we first decompose the frequency-dependent radiation density $E_\nu(\mathbf{x}, t, \nu)$ as $E_\nu(\mathbf{x}, t, \nu) = \tilde{E}(\mathbf{x}, t) \chi_E(\nu)$, where $\chi_E(\nu)$ is the spectral energy distribution (SED) and it's different for different kinds of ionizing sources. Then we define the grey radiation energy density $E(\mathbf{x}, t)$ as

$$E(\mathbf{x}, t) = \int_{\nu_1}^{\infty} E_\nu(\mathbf{x}, t, \nu) \, d\nu = \tilde{E}(\mathbf{x}, t) \int_{\nu_1}^{\infty} \chi_E(\nu) \, d\nu, \qquad (2.10)$$

Then we define the photoheating rates G and photoionization rates Γ_i^{ph} as (Osterbrock, 1989),

$$G = \frac{cE}{\rho_b} \sum_i^{N_s} \mathbf{n}_i \left[\int_{\nu_i}^{\infty} \sigma_i(\nu) \chi_E(\nu) \left(1 - \frac{\nu_i}{\nu} \right) d\nu \right] \bigg/ \left[\int_{\nu_1}^{\infty} \chi_E(\nu) \, d\nu \right], \qquad (2.11)$$

$$\Gamma_i^{ph} = \frac{cE}{h} \left[\int_{\nu_i}^{\infty} \frac{\sigma_i(\nu) \chi_E(\nu)}{\nu} \, d\nu \right] \bigg/ \left[\int_{\nu_1}^{\infty} \chi_E(\nu) \, d\nu \right], \qquad (2.12)$$

where $\sigma_i(\nu)$ is the ionization cross section for the species $\mathbf{n}_i$, h is Planck's constant,

and ν_i is the frequency ionization threshold for species $\mathtt{n}_i$ ($h\nu_{\mathrm{HI}} = 13.6$ eV, $h\nu_{\mathrm{HeI}} = 24.6$ eV, $h\nu_{\mathrm{HeII}} = 54.4$ eV).

The opacity κ is defined as,

$$\kappa = \sum_{i=1}^{N_s} \mathtt{n}_i \left[\int_{\nu_i}^{\infty} \sigma_i(\nu)\chi_E(\nu)\,\mathrm{d}\nu \right] \Big/ \left[\int_{\nu_i}^{\infty} \chi_E(\nu)\,\mathrm{d}\nu \right]. \qquad (2.13)$$

η is the emissivity of ionizing sources. The way we assign it differs from simulation to simulation, so we will discuss it separately in later chapters.

We call this way of radiation transfer the "grey" FLD approximation because we integrate over the frequency-dependent radiation density $E_\nu(\mathbf{x}, t, \nu)$ and approximate it by one field $E(\mathbf{x}, t)$ (Eqn. (2.10)). It is better than the monochromatic radiation transfer because it could include the contributions of photons with a range of energy, but it's not as accurate as multigroup FLD radiation transfer in that it could not represent the spectral shape of the radiation field and the preionization ahead of the I-front (Norman et al., 2015).

2.2.2 Multigroup FLD

For simulations in Chapter 4, we use the multigroup flux-limited diffusion (FLD) approximation of radiation transport. Instead of approximating frequency-dependent radiation density $E_\nu(\mathbf{x}, t, \nu)$ with one radiation energy field like what we did in grey FLD, here we approximate $E_\nu(\mathbf{x}, t, \nu)$ with multiple groups of radiation energy densities and evolve them separately. We consider N_f radiation fields with frequency groups

$$g_1 = [\nu_{1,L}, \nu_{1,R}), \qquad g_2 = [\nu_{2,L}, \nu_{2,R}), \qquad \cdots \qquad g_{N_f} = [\nu_{N_f,L}, \nu_{N_f,R}). \qquad (2.14)$$

We then assume that the frequency-dependent radiation density $E_\nu(\mathbf{x}, t, \nu)$ is piecewise constant:

$$E_\nu(\mathbf{x}, t, \nu) = \begin{cases} \frac{1}{|g_\omega|} E_\omega(\mathbf{x}, t), & \text{if} \quad \nu \in g_\omega, \ \omega = 1, \ldots, N_f \\ \\ 0, & \text{otherwise,} \end{cases} \tag{2.15}$$

where $|g_\omega| = \nu_{\omega,R} - \nu_{\omega,L}$ is the width of the frequency group in Hz, $E_\omega(\mathbf{x}, t) = \int_{\nu_{\omega,L}}^{\nu_{\omega,R}} E_\nu(\mathbf{x}, t, \nu) \, d\nu$ is in unit of erg/cm^3, while $E_\nu(\mathbf{x}, t, \nu)$ is in unit of erg/cm^3/Hz.

The multi-group radiation energy density equations we solve are

$$\partial_t E_\omega - \nabla \cdot (D_\omega \nabla E_\omega) + \frac{3\dot{a}}{a} E_\omega - (\Diamond) = \eta_\omega - c\kappa_\omega E_\omega, \quad \omega = 0, \ldots, N_f - 1, \tag{2.16}$$

where $(\Diamond)$ corresponds to the frequency coupling term:

$$\int_{\nu_{\omega,L}}^{\nu_{\omega,R}} \frac{\nu \dot{a}}{a} \partial_\nu E_\nu \, d\nu \approx \frac{\dot{a} \, \nu_{\omega,R}}{2a} \left(\frac{E_\omega}{|g_\omega|} + \frac{E_{\omega+1}}{|g_{\omega+1}|} \right) - \frac{\dot{a} \, \nu_{\omega,L}}{2a} \left(\frac{E_{\omega-1}}{|g_{\omega-1}|} + \frac{E_\omega}{|g_\omega|} \right) - \frac{\dot{a}}{a} E_\omega.$$

$$\tag{2.17}$$

The multi-group emissivities η_ω and opacities κ_ω, $\omega = 1, \ldots, N_f$, are defined as

$$\eta_\omega = \int_{\nu_{\omega,L}}^{\nu_{\omega,R}} \eta_\nu \, d\nu, \tag{2.18}$$

$$\kappa_\omega = \frac{1}{|g_\omega|} \int_{\nu_{\omega,L}}^{\nu_{\omega,R}} \sum_{i=1}^{N_{\text{chem}}} \sigma_i(\nu) \, \mathbf{n}_i \, d\nu = \sum_{i=1}^{N_{\text{chem}}} \mathbf{n}_i \left(\frac{1}{|g_\omega|} \int_{\nu_{\omega,L}}^{\nu_{\omega,R}} \sigma_i(\nu) \, d\nu \right), \tag{2.19}$$

respectively. Here $\sigma_i(\nu)$ is the cross-section for species i, $\mathbf{n}_i$ is the number density of species i. For more information about the η_ν we choose and the way we place non-zero emissivities into our simulations, please see the source model descriptions in Chapter 4 for details.

Furthermore, to couple the radiation to the matter, we define the photoioniza-

tion rates Γ_i^{ph} and the photo-heating rate $G(\mathbf{x}, t)$. For each species i, the photoioniza-

tion rate Γ_i^{ph} is:

$$
\Gamma_i^{ph}(\mathbf{x}, t) = c \int_{\bar{\nu}_i}^{\infty} \frac{\sigma_{\mathbf{n}_i}(\nu) E_\nu(\mathbf{x}, t, \nu)}{h\nu} \, \mathrm{d}\nu = c \sum_{\omega=1}^{N_f} \int_{\nu_{\omega,L}}^{\nu_{\omega,R}} \frac{\sigma_{\mathbf{n}_i}(\nu) E_\omega(\mathbf{x}, t)}{h\nu |g_\omega|} \, \mathrm{d}\nu \qquad (2.20)
$$

$$
= \sum_{\omega=1}^{N_f} \frac{c E_\omega(\mathbf{x}, t)}{h} \left(\frac{1}{|g_\omega|} \int_{\nu_{\omega,L}}^{\nu_{\omega,R}} \frac{\sigma_{\mathbf{n}_i}(\nu)}{\nu} \, \mathrm{d}\nu \right), \qquad i = 1, \ldots, N_{\mathrm{chem}},
$$

where we assume that $\sigma_{\mathbf{n}_i}(\nu) = 0$ for all $\nu < \bar{\nu}_i$. We sum over species to get one

photo-heating rate $G(\mathbf{x}, t)$:

$$
G(\mathbf{x}, t) = \frac{c}{\rho_b} \sum_{i=1}^{N_{\mathrm{chem}}} \mathbf{n}_i \int_{\bar{\nu}_i}^{\infty} \sigma_{\mathbf{n}_i}(\nu) E_\nu(\mathbf{x}, t, \nu) \left(\frac{h\nu - h\bar{\nu}_i}{h\nu} \right) \, \mathrm{d}\nu
$$

$$
= \sum_{i=1}^{N_{\mathrm{chem}}} \sum_{\omega=1}^{N_f} \frac{c \, \mathbf{n}_i \, E_\omega(\mathbf{x}, t)}{\rho_b} \left(\frac{1}{|g_\omega|} \int_{\nu_{\omega,L}}^{\nu_{\omega,R}} \sigma_{\mathbf{n}_i}(\nu) \left(1 - \frac{\bar{\nu}_i}{\nu} \right) \, \mathrm{d}\nu \right). \qquad (2.21)
$$

2.3 Numerical methods

2.3.1 The Enzo code

We run all our simulations using the publicly available code Enzo (O'Shea et al.,

2004; Norman et al., 2007; Bryan et al., 2014a). Enzo uses an N-body particle-mesh

(PM) solver to calculate the movement of collisionless dark matter particles. We use

the Cloud-In-Cell (CIC) formulation with leapfrog time integration, which is formally

second order-accurate in space and time. For fields like mass density we use the

piecewise parabolic method (PPM) (Colella and Woodward, 1984) to evolve them. We

assume that all chemical species have the same velocity. PPM is also formally second

order accurate in space and time. We calculate the gravitational potential by solving

the Poisson equation on the uniform grid using 3-dimensional fast Fourier transforms (FFT). Enzo is also capable to do simulations with adaptive mesh refinement (AMR) for gas dynamics, but AMR is not used in this work.

In Enzo we could choose to run with 6-, 9-, or 12-chemical species models. In this work we use 6-species model, including H I, H II, He I, He II, He III and the electron e^-. To calculate radiative heating and cooling, Enzo includes atomic line excitation, collisional excitation, recombination, free-free transitions, molecular line cooling, Compton scattering of the cosmic microwave background, and different kinds of ultraviolet background that heats the gas by photoionization and/or photodissociation.

2.3.2 Operator split solution procedure

We calculate all equations described in Section 2.2 in Enzo. Here all our simulations are run with unigrid, so the simulation space is discretized using a uniform grid. Simulation time is discretized using an operator split time-stepping method, where we separate all equations we solve into 5 components.

The first component is the self-gravity equation (2.1),

$$\nabla^2 \phi = \frac{4\pi g}{a}(\rho_b + \rho_{dm} - \langle \rho \rangle), \tag{2.22}$$

which solves for the gravitational potential ϕ. We solve it using our own 3D Fast Fourier Transform solver. To solve ϕ we need baryon density field ρ_b and dark matter density field ρ_{dm}. ρ_b is defined as a gridded Eulerian field, whereas ρ_{dm} is computed from dark matter particles' positions $\mathbf{x}_i^n$ using the Cloud In Cell (CIC) mass assignment algorithm (Hockney and Eastwood, 1988).

The second component includes the equations of conservation,

$$\partial_t \rho_b + \frac{1}{a}\mathbf{v}_b \cdot \nabla \rho_b = -\frac{1}{a}\rho_b \nabla \cdot \mathbf{v}_b,$$

$$\partial_t \mathbf{v}_b + \frac{1}{a}\left(\mathbf{v}_b \cdot \nabla\right)\mathbf{v}_b = -\frac{\dot{a}}{a}\mathbf{v}_b - \frac{1}{a\rho_b}\nabla p - \frac{1}{a}\nabla \phi,$$

$$\partial_t e + \frac{1}{a}\mathbf{v}_b \cdot \nabla e = -\frac{2\dot{a}}{a}e - \frac{1}{a\rho_b}\nabla \cdot (p\mathbf{v}_b) - \frac{1}{a}\mathbf{v}_b \cdot \nabla \phi, \qquad (2.23)$$

$$\partial_t E + \frac{1}{a}\nabla \cdot (E\mathbf{v}_b) = 0,$$

$$\partial_t \mathbf{n}_i + \frac{1}{a}\nabla \cdot (\mathbf{n}_i\mathbf{v}_b) = 0, \qquad i = 1, \ldots, N_s.$$

Note that the above equations for E and $\mathbf{n}_i$ do not include emissivities, recombination or photoionization, which are included in later components. We solve these equations explicitly using the Piecewise Parabolic Method (PPM) (Colella and Woodward, 1984).

The third component corresponds to the grey radiation energy equation (2.8),

$$\partial_t E = \nabla \cdot (D\nabla E) - \frac{\dot{a}}{a}E - c\kappa E + \eta. \qquad (2.24)$$

The algorithm to solve this equation is described in detail in Reynolds et al. (2009). Since the characteristic time scale for radiation transport is much faster than hydrodynamic motion, we use an implicit θ-method for time discretization, allowing both backwards Euler and trapezoidal implicit quadrature formulas. With implicit FLD approximation for the radiation transport, we solve for the change in the radiation energy field, $\delta E = E^{n+1} - E^n$, in linearized equations over each time step. We use hypre library to solve them, which has optimal parallel scalability to supercomputer architectures.

The fourth component includes photoheating, gas cooling, recombination, and

photoionization processes,

$$\partial_t e = G - \Lambda, \tag{2.25}$$

$$\partial_t \mathbf{n}_i = \alpha_{i,j} \mathbf{n}_e \mathbf{n}_j - \mathbf{n}_i \Gamma_i^{ph}, \qquad i = 1, \ldots, N_s.$$

Since the time scales of these processes are close to those of third component, we also solve them implicitly in time. We build two different solvers to solve them. The first solver is more loosely coupled, using a single Jacobi iteration of a linearly-implicit backwards Euler discretization for each species in each cell. It does not attempt to accurately resolve the nonlinearity in these equations, nor does it iterate between the different species in each cell to achieve a fully self-consistent solution, but its results are typically accurate to within 10% relative error, and it's relatively more efficient.

The second solver approximates the equations using an implicit quasi-steady-state formulation, in which we assume a fixed ionization state $(\mathbf{n}_i^{n-1} + \mathbf{n}_i^n)/2$, and a fixed gas energy $(e^{n-1} + e^n)/2$ when evolving the time step from t^{n-1} to t^n. We then solve the resulting set of differential equations analytically to obtain e^n and $\mathbf{n}_i^n$. Since the coupling between the gas energy and chemical ionization is much tighter, this solver is more expensive, but may result in a more accurate and stable solution than the first solver.

The fifth component corresponds the calculation or assignment of emissivity fields. If we turn on the star formation and feedback processes, the emissivity field would be evaluated, and the density field and gas energy field would be updated by

integrating the equations

$$\partial_t \rho_b = -\dot{\rho}_{SF}, \tag{2.26}$$

$$\partial_t e = \dot{e}_{SF}. \tag{2.27}$$

In our simulations here we turn off the star formation and feedback processes, and assign the emissivity fields either directly through inline python (Chapter 7) or through MGFLD (Chapter 4). In these simulations we don't need to update ρ_b and e through Eqn. (2.26) and (2.27). Please see relevant chapters for more details.

2.3.3 Radiation subcycling

Since the radiation (Eqn. (2.24)) and gas energy/chemistry (Eqn. (2.25)) components evolve at comparable time scales that are both much faster than the hydrodynamic motions, we apply an adaptive time-stepping strategy, where the radiation component limits the overall time step, choosing a conservative time step to ensure consistency among all physical processes. The time step estimation algorithm is the same as in Reynolds et al. (2009).

When using the loosely-coupled solver for the fourth component, the order of these processes from t^{n-1} to t^n is:

```
Set  t_hydro = t_chem = t_rad = t_dm = t^(n-1).
Set  Δt = min{Δt_hydro, Δt_expansion, Δt_rad},  and  t^n = t^(n-1) + Δt.
While  (t_rad < t^n)

     Try to evolve the  E(t)  according to Eqn.(2.24).
     If failure, set  Δt_rad = 0.1Δt_rad.
     Else set  t_rad = t_rad+Δt_rad  and update  Δt_rad  based on accuracy
     estimates.
```

```
Post-process E(tⁿ) to compute G and Γᵢᵖʰ.
Compute φ using Eqn.(2.22), and post-process to generate ∇φ.
Evolve the hydrodynamics sub-system Eqn.(2.23), t_hydro → t_hydro +
Δt.
While (t_chem < tⁿ)

    Set Δt_chem based on accuracy estimates.
    Evolve the chemical and gas energy subsystem Eqn.(2.25),
    t_chem → t_chem + Δt_chem.

Evolve the dark matter particles, t_dm → t_dm + Δt.
Compute η from star formation or from inline-python/MGFLD.
```

When using the tightly-coupled solver for the fourth component, the order

changes slightly:

```
Set t_hydro = t_chem = t_rad = t_dm = tⁿ⁻¹.
Set Δt = min{Δt_hydro, Δt_expansion, Δt_rad}.
While (t_rad < tⁿ)

    Try to evolve the radiation field according to Eqn.(2.24).
    If failure, set Δt_rad = 0.1 * Δt_rad.
    Else
        Set t_rad = t_rad + Δt_rad and update Δt_rad based on
        accuracy estimates.
        Post-process E(t_rad) to compute G and Γᵢᵖʰ.
        While (t_chem < t_rad)
            Set Δt_chem based on accuracy estimates.
            Evolve the chemical/energy subsystem Eqn.(2.25),
            t_chem → t_chem + Δt_chem.

Compute φ using Eqn.(2.22), and post-process to generate ∇φ.
Evolve the hydrodynamics sub-system Eqn.(2.23),
t_hydro → t_hydro + Δt.
Evolve the dark matter particles, t_dm → t_dm + Δt.
Compute η from star formation or from inline-python/MGFLD.
```

2.4   Tests

In this section we show two test results: (1) H I reionization tests. Reynolds

et al. (2009) and Norman et al. (2015) performed a H I reionization test problem and

compared with analytical solutions to verify the new coupling strategy between the radiation transport and chemistry in Enzo. Here we briefly show their results. (2) He II reionization tests. To verify our new MGFLD algorithm added in Enzo, we reproduced He II reionization test problem described in McQuinn et al. (2009) and compared with their results.

2.4.1 H I reionization test

To verify the updated coupling strategy between the radiation transport and chemistry, Reynolds et al. (2009) did an isothermal ionization of a static neutral hydrogen (H I) region using Enzo. This test problem is similar with the Test 1 in Iliev et al. (2006). They both simulate the expansion of an ionized hydrogen (H II) bubble in the uniform H I gas containing an ionizing source. They also assume that initially the gas is static and has a uniform temperature. With these assumptions, the radius of the ionized bubble should first increase then reach a final balanced value when ionizations are balanced with recombinations. The radius of the ionization front $r(t)$ can be calculated analytically:

$$r(t) = r_s \left(1 - e^{-t/t_{rec}}\right)^{1/3}, \quad \text{where} \quad r_s = \left(\frac{3\,\dot{N}_\gamma}{4\pi\,\alpha_B\,n_H^2}\right)^{1/3}. \tag{2.28}$$

Here, r_s is the final balanced radius, called the Strömgren radius. $t_{rec} = (\alpha_B\,n_H)^{-1}$ is the recombination time, where α_B is the case B hydrogen recombination rate. $\dot{N}_\gamma$ is the emission rate of ionizing photons. n_H is the hydrogen number density of the gas.

In their tests, they set $\dot{N}_\gamma = 5 \times 10^{48}$ photons s^{-1}, $n_H = 10^{-3}$ cm^{-3}, $\alpha_B = 2.59 \times 10^{-12}$ cm^{2}s^{-1}, domain $[0, 6.6\,\text{kpc}]^3$, temperature $T = 10^4$ K, and time interval
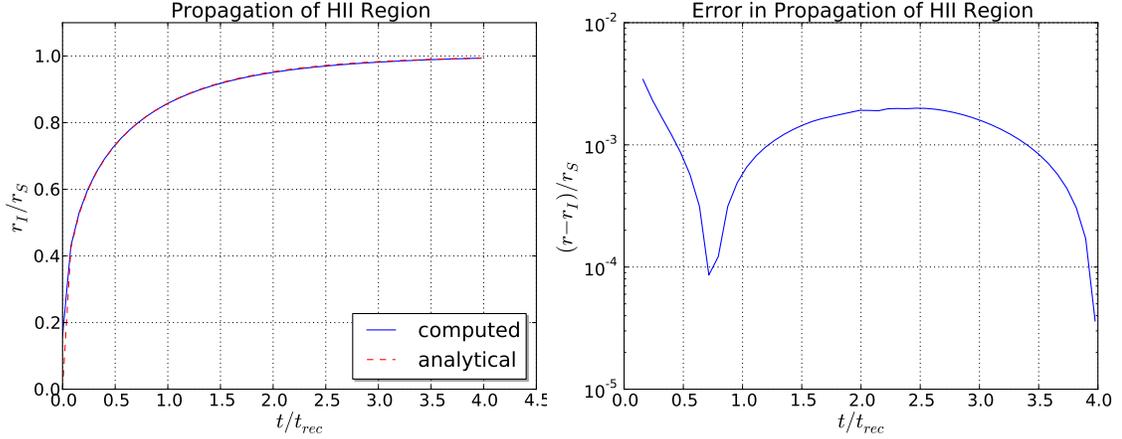
Figure 2.1: Left: Comparison between computed and analytical ionization front position for the isothermal ionization test using a 128^3 mesh and time step tolerance $\tau_{tol} = 10^{-4}$. Right: The error in the ionization front position versus time as defined in Eqn. (2.29).

$[0, 5\,\text{Myr}]$. The ionization source is assumed to be monochromatic, at the H I ionization frequency $h\nu = 13.6$ eV, and is located at the location $(0, 0, 0)$. For initial conditions, they set $E = 10^{-45}$ erg cm^{-3} and ionization fraction H II/H $= 0.0012$. They use reflecting boundary conditions for the radiation field at the $x = 0$, $y = 0$, and $z = 0$ faces, and outflow boundary conditions at the other three faces.

To compare the computed and analytical radius of ionization front numerically, they define the error in the ionization front position as

$$error \;=\; \left\|\frac{r_{computed} - r_{true}}{r_s}\right\|_{RMS} \;=\; \left(\frac{1}{N_t}\sum_{i=1}^{N_t}\left(\frac{r_{computed,i} - r_{true,i}}{r_s}\right)^2\right)^{1/2}. \quad (2.29)$$

Fig. 2.1 shows the computed and analytical radius of ionization front, and the error versus time, where they use a 128^3 spatial grid and time step tolerance $\tau_{tol} = 10^{-4}$.

Figure 2.2 shows slices through the simulation box, to verify the convergence of the size of the ionized region with different spatial resolutions.
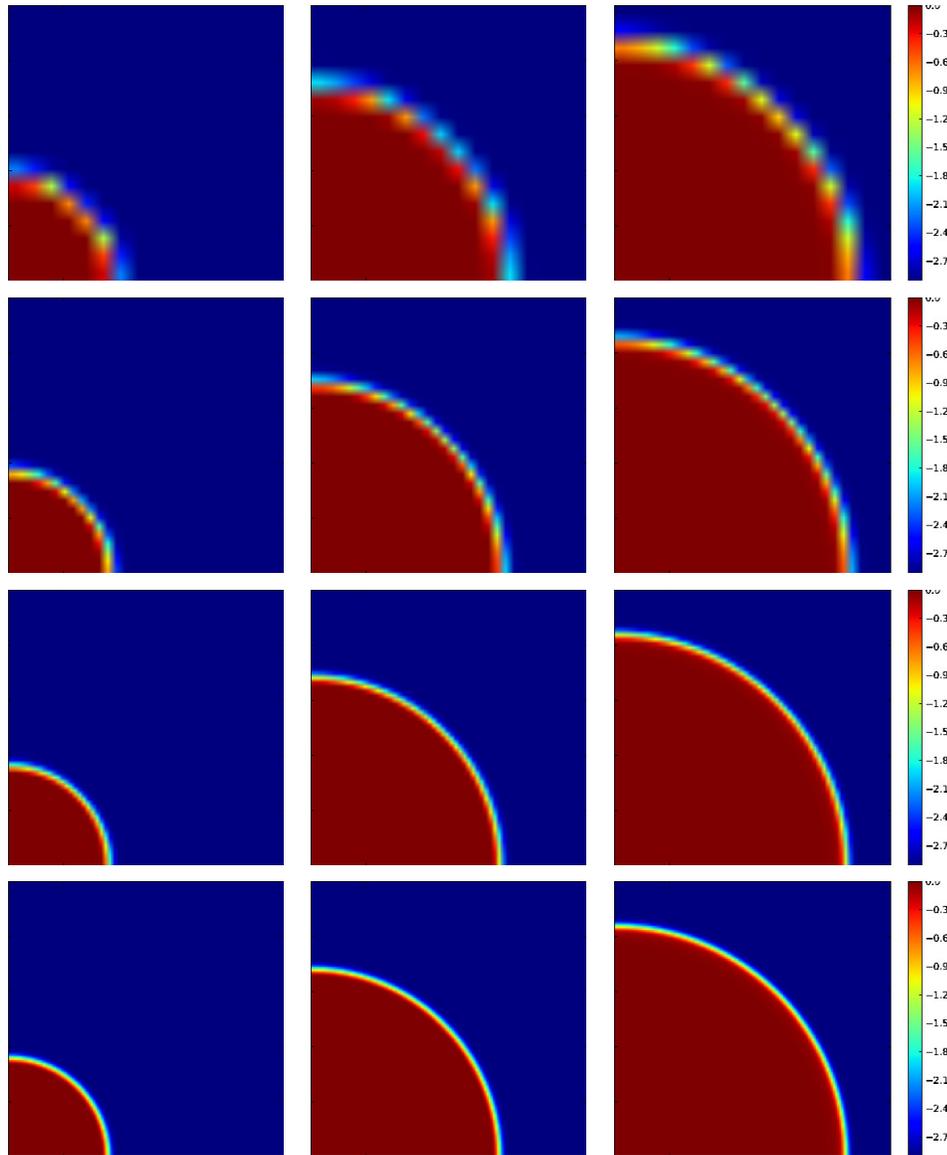
Figure 2.2: Slices perpendicular to the z axis of the simulation box of H I reionization tests. Colorbars to the right label the ionized fraction (H II/(H I + H II)) in log scale. From left to right it shows the evolution of the ionized region at 10, 100 and 500 Myr, respectively. From top to bottom it shows simulations with spatial grids of 16^3, 32^3, 64^3 and 128^3, respectively. It demonstrates the convergence of results with different spatial resolutions.

2.4.2   He II reionization test

In this section we conduct three idealized He II reionization simulations to test the newly developed MGFLD algorithm. Two simulations have L_{box}=100 Mpc, and grid sizes of 128^3 and 256^3. A third has L_{box}=50 Mpc, and a grid size of 256^3. These three are chosen to vary the cell resolution by a factor of 4 and span the range of resolutions used in our production simulations presented in Chapter 4.

Initially we set the density equal to the mean density at $z = 4$ and set $x_{\mathrm{HII}} = 1$, $x_{\mathrm{HeII}} = 1$. A source with the spectral index $\alpha_{\mathrm{UV}} = 1.5$ and intensity $\dot{N} = 10^{54}$ He II ionizing photons per second is put at the center of the simulation box. We use our MGFLD algorithm with 5 energy bins ranging from 54.4eV to 400eV to transfer the radiation. Specifically, the energy bins are [54.4 eV, 65 eV], [65 eV, 75 eV], [75 eV, 125 eV], [125 eV, 155 eV], and [155 eV, 400 eV]. McQuinn et al. (2009) has shown that these 5 energy bins are sufficient to converge on the temperature and ionization structure of this test problem. The volume weighted temperature and He III fraction after 50 Myr and 250 Myr are shown in Fig. 2.3. It shows the expansion of the ionized bubble. Consistent results are reached with different resolutions, indicating good convergence. Behind the ionization front, at 50 Myr the gas temperature is raised from 3000K to 14,000K by He II photoionization and heating, an increase of $\sim 11,000$K in good agreement with the results of McQuinn et al. (2009). At 250 Myr, the peak temperature has increased slightly to 18,000K. The temperature distribution interior to the I-front increases linearly from a central temperature of $\sim 10,000$K to the peak temperature. We note that the ionization front is not sharp, but rather smooth. Our convergence study shows that this is not due to numerical broadening. Rather, it is due to pre-heating and pre-ionization of the gas ahead of the I-front by
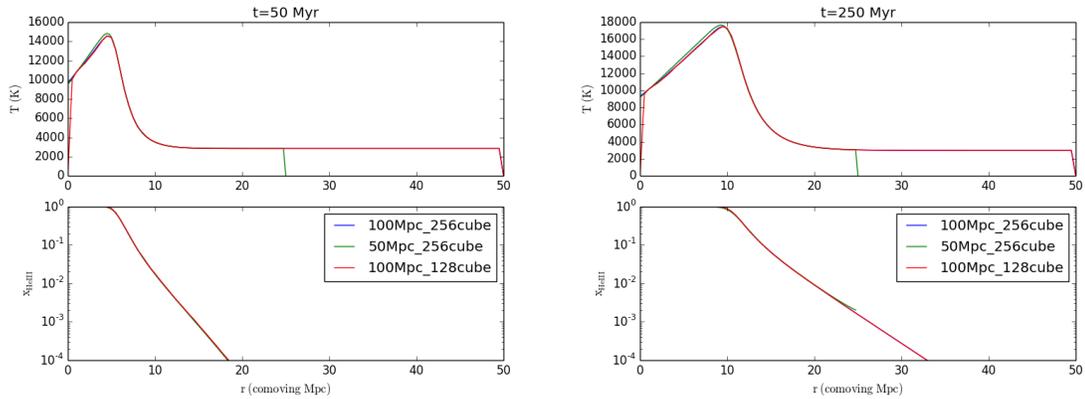
Figure 2.3: Volume weighted temperature (upper panel) and He III fraction (lower panel) at 50 Myr (Left) and 250 Myr (Right) after $z = 4$ from three test problems: a) L_{box}=100Mpc, 256^3, b) L_{box}=50Mpc, 256^3 and c) L_{box}=100Mpc, 128^3.

the higher energy, more penetrating photons (Iliev et al., 2006). We see that as time goes on, the I-front broadens considerably.

Chapter 3

Standard Lyman Alpha Forest Simulations

3.1 Simulations

In this chapter we present several simulations performed with Enzo to examine the mismatch between the standard LAF simulations and observations, as discussed in Section 1.4. The output of the simulations was analyzed with the software yt (Turk et al., 2011a) using scripts described in the Software appendix. In all simulations here we use the UV background from Haardt and Madau (2012) (HM2012), but with several different modifications to examine how the IGM temperature evolves under different assumptions. The optically thin photoionization rates $\Gamma(z)$ and photoheating rates $\mathscr{H}(z)$ of H I, He I, and He II from HM2012 are shown in Fig. 3.1. Most photons ionizing H I and He I in IGM were emitted by early star-forming galaxies when $z > 6$ (Becker et al., 2001; Fan et al., 2005), while most He II was ionized later by quasars (Madau and Meiksin, 1994; Miralda-Escudé et al., 2000).
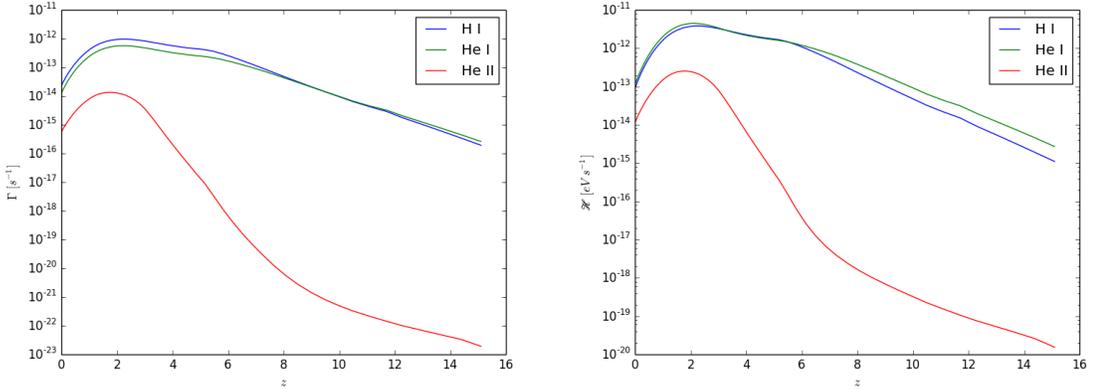
Figure 3.1: Photo-ionization rates (Left) and photo-heating rates (Right) from Haardt and Madau (2012) (HM2012) for H I, He I, and He II versus redshift. Those backgrounds are modified and applied to our simulations (Table 3.1).

We examine the thermal evolution of the IGM under different assumptions using 4 simulations with parameters shown in Table 3.1. In the first three, we vary the amount of heating due to He II reionization in a scenario where we assume H I reionization is complete at $z = 6.5$. In the fourth we simulate a larger volume with HM2012 UVB active over the full redshift range where it is calculated (Haardt and Madau, 2012). In simulation 1024, we use the full HM2012 background from $z = 6.5$ to $z = 2$. We also include a redshift-dependent Compton heating term. To study how the photoionization and photoheating of He II contributes to the thermal history of IGM, we did simulation 1024_0, where we removed the photoionization rates and photoheating rates for He II, with other things being the same as simulation 1024. Since all simulations here are optically thin, i.e. no radiation transfer is used, we multiply the He II photoheating rates by a factor X_{228} to account for the extra heating that might be experienced by the optically-thick gas (Abel and Hähnelt, 1999; Jena et al., 2005). In simulation 1024_1.8 we set $X_{228} = 1.8$, with other parameters being the same as simulation 1024. We then did simulation 2048, which has a box size

twice as big as simulation 1024, but with the same cell size. The grid resolution of all simulations is 19.5 comoving kpc/h which Lukić et al. (2015) have demonstrated is sufficient to converge on the synthetic Lyman alpha forest's observable properties to 1% accuracy. In all simulations we've turned off all star formation and their feedback effects.

Table 3.1: List of standard LAF simulations.

Name	Grids	Box size [Mpc/h]	Background	X_{228}	z_{BG}
1024_0	1024^3	20	HM2012 without He II reionization	0	6.5
1024	1024^3	20	HM2012 with He II reionization	1.0	6.5
1024_1.8	1024^3	20	HM2012 with He II reionization	1.8	6.5
2048	2048^3	40	HM2012 with He II reionization	1.0	15.1

Notes: HM2012 refers to the UV background from Haardt and Madau (2012). z_{BG} is the redshift that the UV background is turned on.

For all simulations in this chapter, we use the TT,TE,EE+lowP+lensing+ext set of cosmological parameters from the Planck Collaboration et al. (2015): $\Omega_b = 0.0487$, $\Omega_m = 0.309$, $\Omega_\Lambda = 0.691$, $H_0 = 67.7$km/s/Mpc, $\sigma_8 = 0.816$, $n_s = 0.967$.

3.2   General results

Before performing the simulations shown in Table 3.1, we did a test simulation with 256^3 grids and the same random seed. We ran it to $z = 4$ and found the location of the largest halo at that time. Then, for all later simulations, we re-centered them to that location. That is why a large halo is in the center of the box. Fig. 3.2 shows plots of simulation 1024. All slices (top 4 figures) are parallel to x-y plane and at axis z=0.5 in simulation unit. The slices show the standard filament-void structure of the IGM
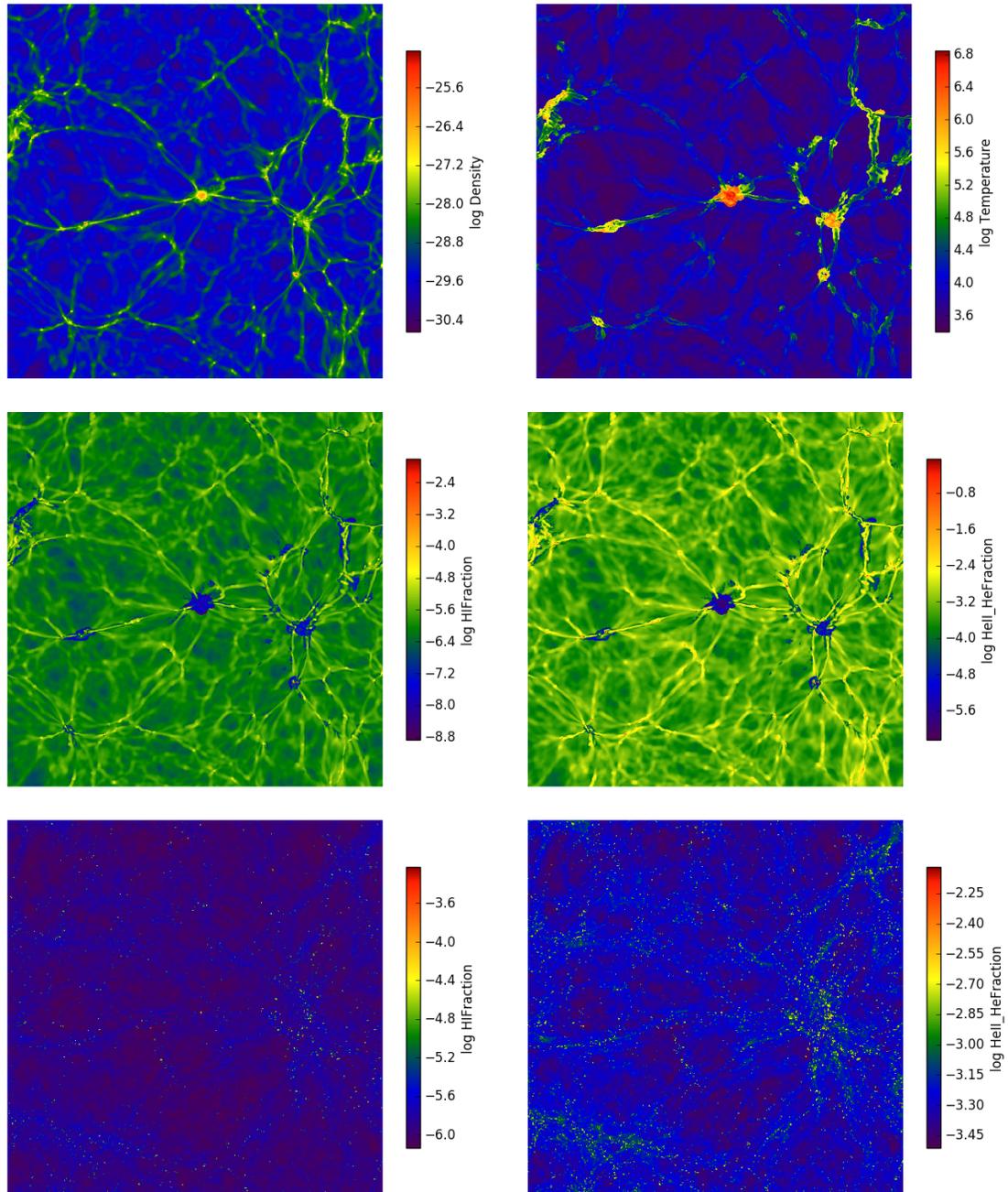
Figure 3.2: Slices and projections of simulation 1024 at $z = 2$. Top Left: slices of baryon density. Top Right: slices of temperature. Middle Left: slices of H I fraction. Middle Right: slices of He II fraction. Bottom Left: projections of H I fraction. Bottom Right: projections of He II fraction. All slices are made perpendicular to z axis at z=0.5 in simulation unit.

induced by the gravitational clustering of the cold dark matter in the simulation (not shown). The filaments are the origin of Lyman alpha forest absorption, as discussed in Chapter 1. They also show that, typically, a region with higher density would have a higher temperature, a lower H I fraction (defined by $\rho_{\text{H I}}/(\rho_{\text{H I}} + \rho_{\text{H II}})$), and a lower He II fraction (defined by $\rho_{\text{He II}}/(\rho_{\text{He I}} + \rho_{\text{He II}} + \rho_{\text{He III}})$). This is due to the large structure formation at those locations. However, the projections of H I (bottom left) and He II (bottom right) do show dot-like regions with relatively high fractions of H I and/or He II. Most of these regions correspond to dense gas in large halos, and might serve as absorbers for H I and/or He II Lyα lines.

It has been known since the early days of numerical simulations of the Lyman alpha forest that the density and temperature of the photoionized IGM is highly correlated. To see this in Fig. 3.3 we show phase plots of baryon overdensity versus temperature for simulation 1024 (Left) and 1024_1.8 (Right) at different redshifts. The temperature of IGM is mostly determined by the balance of photoheating and adiabatic cooling, and it could be approximated by a power law in the baryon density (Hui and Gnedin, 1997):

$$T = T_0\Delta^{\gamma-1}, \tag{3.1}$$

where T_0 is the IGM temperature at the mean density, Δ is the IGM density in units of mean baryon density, and $\gamma - 1$ is the power law index. As shown in the plots, the majority of IGM (i.e. the red and yellow regions) could be approximately described by this equation of state, while other high temperature regions (i.e. the green and blue regions) represent the shock-heated gas. During the He II reionization, we anticipate a roughly constant increase of temperature independent of density by our uniform UV background (Puchwein et al., 2014). For regions with lower (higher)
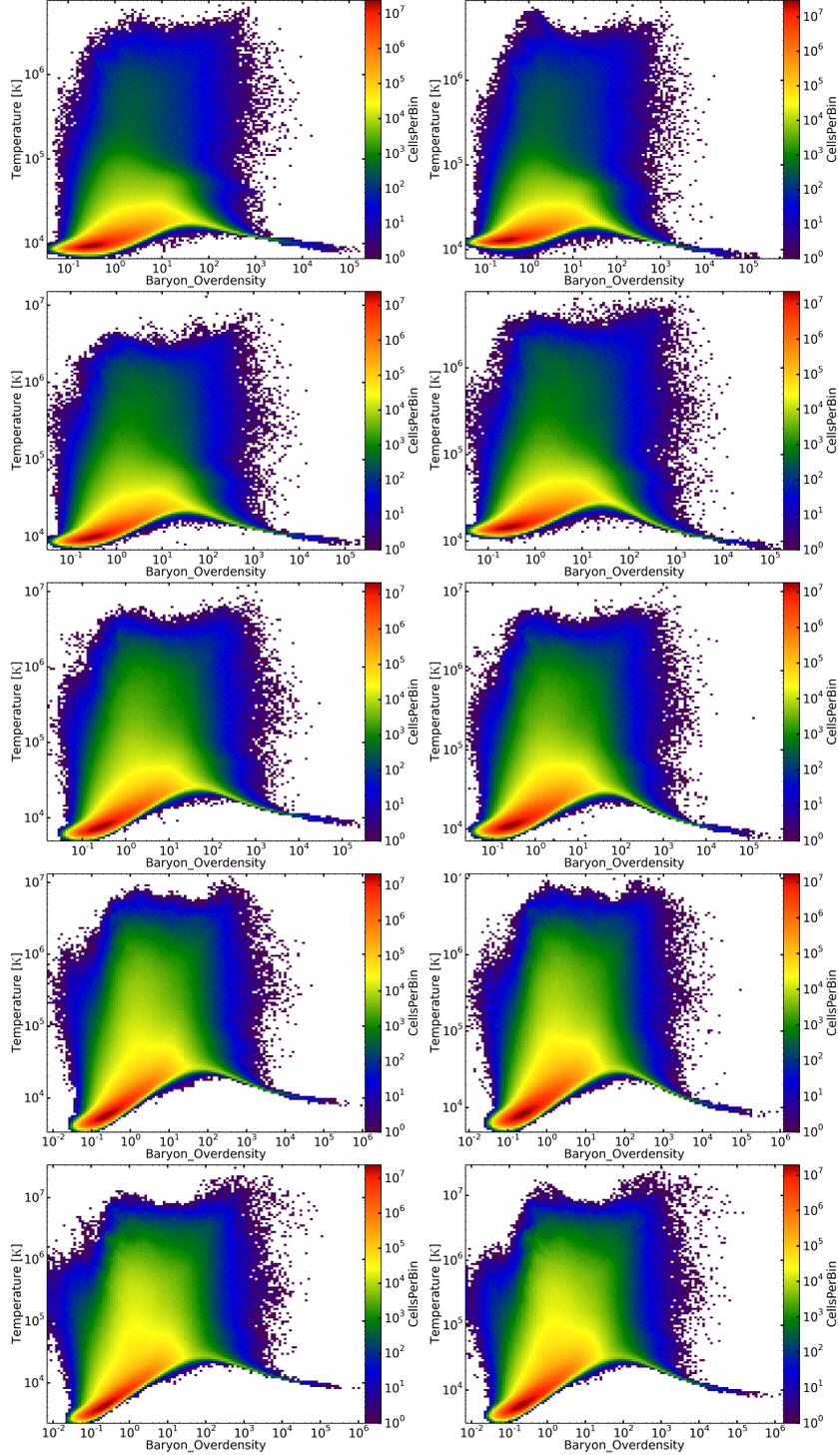
Figure 3.3: Baryon overdensity versus Temperature phase diagrams. Left: simulation 1024. Right: simulation 1024_1.8. From top to bottom, the plots are made at $z = 4.0, 3.5, 3.0, 2.5, 2.0$, respectively. Color bars indicate the count of cells in each 2-dimensional bin.
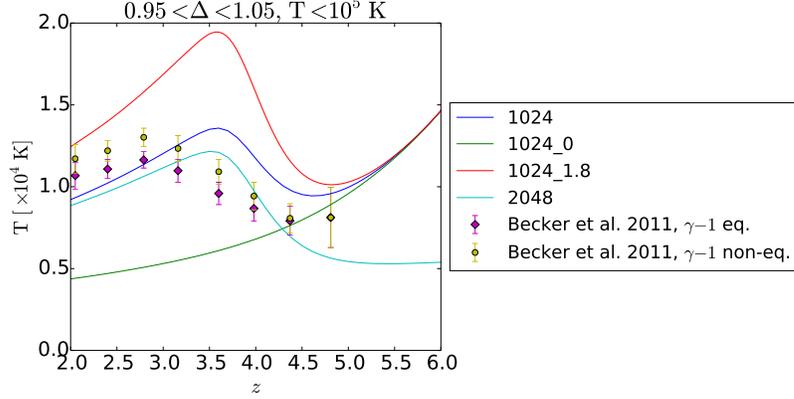
Figure 3.4: Evolution of T_0 for simulations 1024, 1024_1.8 and 1024_0. At each redshift, T_0 is calculated from the average of logarithmic temperature of cells with $0.95 < \Delta < 1.05$ and $T < 10^5$ K.

temperatures before He II reionization, this corresponds to a higher (lower) increase in the temperature in log scale. We thus expect a flatter slope in the $T - \Delta$ relationship in log-log space, as shown in the top rows in Fig. 3.3. Then, after the He II reionization completes, since the denser region cools down more efficiently, the slope increases back, as we see in the bottom rows. The effect of extra photoheating is also visible in Fig. 3.3, since the majority of cells in simulation 1024_1.8 (i.e. the red regions in right panels) have a higher temperature than those in simulation 1024 (i.e. the red regions in left panels).

Fig. 3.4 shows the evolution of the IGM temperature at the mean density T_0 for simulations 1024, 1024_1.8 and 1024_0. We get T_0 from the average of logarithmic temperature of cells with $0.95 < \Delta < 1.05$ and $T < 10^5$ K. It shows how the (extra) photoheating during the He II reionization affects the temperature of IGM. Without reionization of He II (simulation 1024_0), T_0 continues to drop due to adiabatic cooling. With HM2012 UV background including He II photoheating and photoionization (simulation 1024), a bump in temperature appears which peaks at $z \sim 3.5$. With
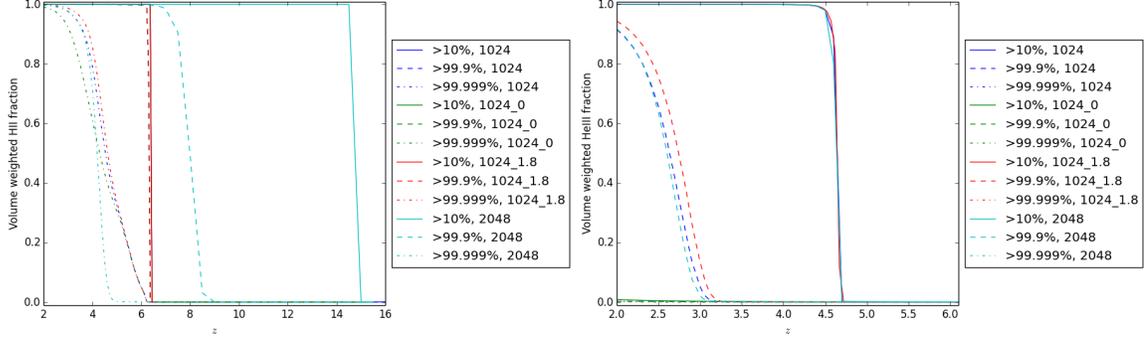
Figure 3.5: Reionization history of H I (Left) and He II (Right) for simulation 1024. The line with legend $> x\%$ indicates the fraction of number of cells with H II (He III) fraction larger than $x\%$.

additional photoheating to He II (simulation 1024_1.8), a higher peak is produced, but its maximum is at the same redshift. We will return to this important point in Chapter 4.

In Fig. 3.5 we show the reionization history of H I (Left) and He II (Right) for simulation 1024. We use different metrics $> x\%$ to measure the degree of ionization. Volume weighted H II (He III) fraction with legend $> x\%$ indicates the fraction of number of cells with H II (He III) ionization fraction larger than $x\%$. It shows that the extra heating of He III, although it increases the IGM temperature, doesn't significantly affect the reionization history of He III. We see that H I becomes highly ionized immediately after z_{BG}. Since observations favor a late reionization scenario (Chapter 7), the $z_{BG} = 6.5$ models are more realistic with respect to the thermal evolution. He II achieves a low level of ionization once the quasar contribution to the UVB becomes non-negligible around $z = 5$ (Fig. 3.1). However He II only becomes highly ionized for $z < 3$ as the QSO background approaches its maximum intensity at $z \sim 2$.
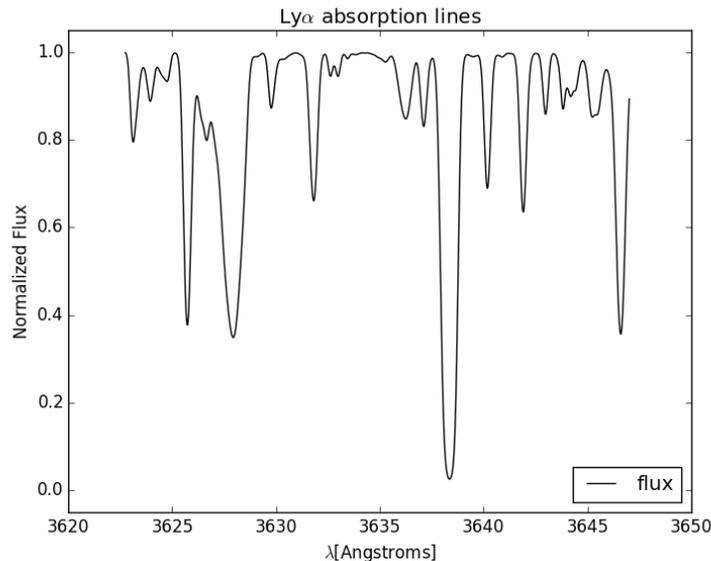
Figure 3.6: A spectrum showing LAF from simulation 1024_1.8. It's generated at $z \sim 2$. The wavelengths shown are observed wavelengths.

3.3 Synthetic H I absorption spectra

Using the light ray generator in yt, we generate 1,000 light rays for each simulation at $z \sim 2$ and $z \sim 3.5$. All light rays have random start positions on the z=0 plane and are parallel to z-axis. Temperature and H I density along each light ray are saved. Then, with the spectrum generator, those values are used to calculate the Voigt profiles. Then the lines are shifted according to the redshift and the line of sight peculiar velocities. Fig. 3.6 shows an example of line profiles (normalized flux versus wavelength) from simulation 1024_1.8.

For each absorption spectrum, we fit the absorption lines with Voigt profiles using the tool AUTOVP (Davé et al., 1997) from which we obtain the column density N and Doppler parameter b for each Voigt profile. Then with all 1,000 spectra for each simulation at a certain redshift, we get the distribution of Doppler parameter b
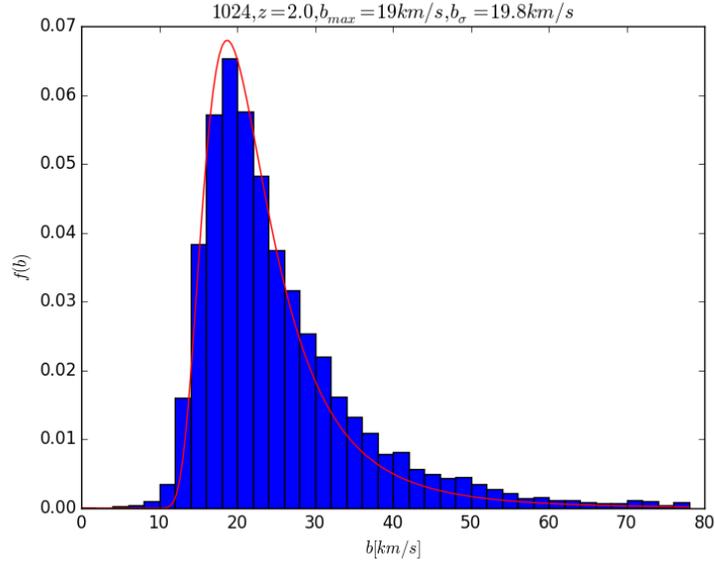
Figure 3.7: The distribution of Doppler parameter b and its fit with Hui-Rutledge function for simulation 1024 at $z \sim 2$. The fitted value of b_σ is 19.8 km/s.

for all absorbers with column density N satisfying $10^{12.5}/\text{cm}^2 < N < 10^{14.5}/\text{cm}^2$, and fit it with the Hui-Rutledge function (Hui and Rutledge, 1999),

$$\frac{dn}{db} = B\frac{b_\sigma^4}{b^5} \exp\left(-\frac{b_\sigma^4}{b^4}\right), \tag{3.2}$$

where $\frac{dn}{db}$ is the number of absorption lines per km/s per unit redshift z. The distribution of b and its fit for simulation 1024 at $z \sim 2$ is shown in Fig. 3.7.

The same analysis is done for $z \sim 3.5$ data dumps. The fitted Hui-Rutledge profiles are shown in Fig. 3.8. The fitted values of b_σ are shown in Table 3.2. It shows that b_σ increases as the photoheating of He II increases (from simulation 1024_0 to 1024 to 1024_1.8) and decreases as time passes by the He II reionization (from $z \sim 3.5$ to $z \sim 2$).
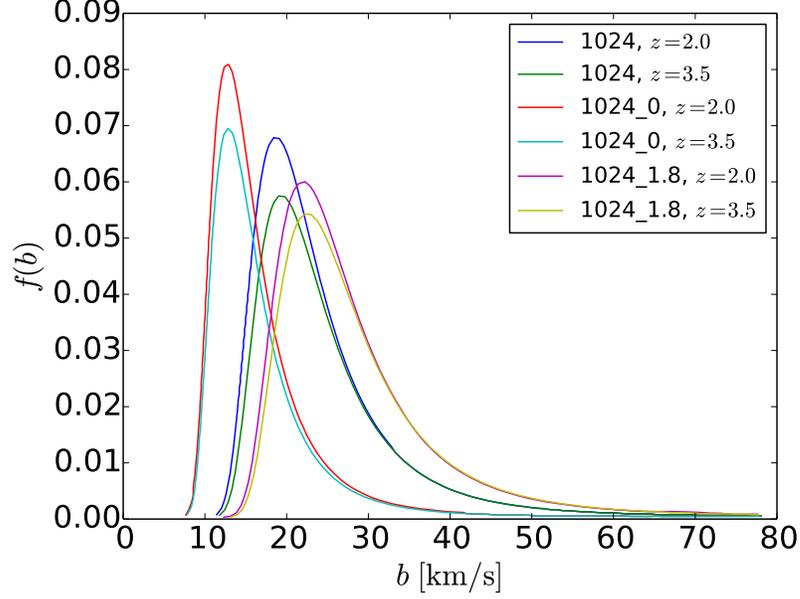
Figure 3.8: The fitted distribution function of b of simulation 1024, 1024_1.8 and 1024_0 at redshift $z \sim 2$ and $z \sim 3.5$.

Table 3.2: Fitted values of b_σ.

b_σ [km/s]	$z \sim 3.5$	$z \sim 2$
1024_0	13.6	13.5
1024	20.5	19.8
1024_1.8	23.9	23.3

Finally note that to resolve the line width and get a reliable estimate of Doppler parameter, we need high resolution for our simulations (Bryan et al., 1999; Tytler et al., 2009). The Hubble broadening of one cell with different cell sizes as a function of redshift is shown in Fig. 3.9. Given the typical magnitude of $b_\sigma \sim 20$ km/s at $2 < z < 4$, the plot shows that 20 kpc/h is good enough to resolve H I LAF line width (Lukić et al., 2015).
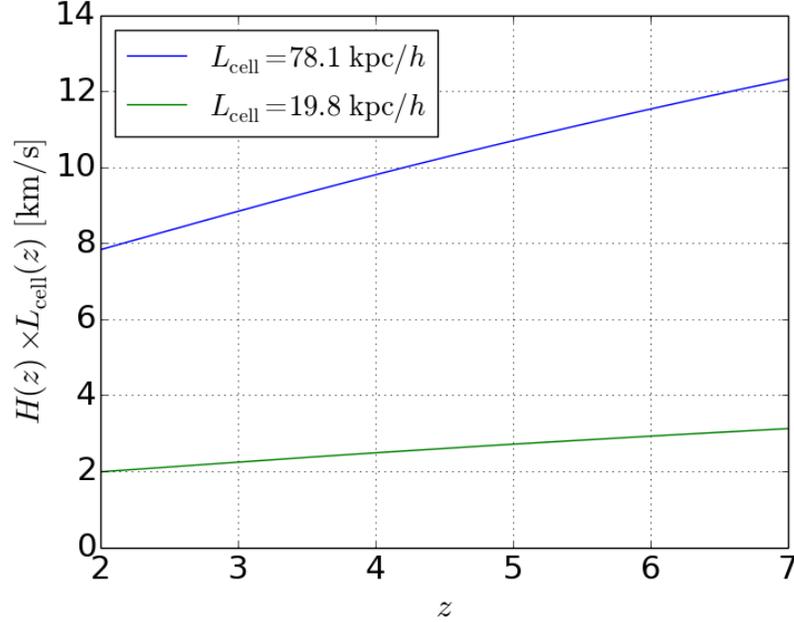
Figure 3.9: The Hubble broadening of one cell with different cell sizes: 19.8 kpc/h, and 78.1 kpc/h as a function of redshift. It shows that in order to resolve LAF at high redshifts, we need higher resolution for simulations.

3.4 Flux PDF

The normalized flux F_i for every pixel ranges from 0 to 1. It can be written as $F_i = e^{-\tau_i}$, where τ_i is the optical depth for that pixel. With all the normalized flux F_i's, we have generated the flux probability distribution function (PDF) $P(F)$. It is the probability density of F so that $\int_0^1 P(F) \, dF = 1$. We then compare our results with the observations in Kim et al. (2007) extrapolated to $z = 2$ and $z = 3$ (Fig. 3.10). Although the shapes of the simulated PDFs are qualitatively similar to the observations, they fall outside the rather stringent error bars. Hence the discrepancy is significant. Specifically, at $z = 2$ it shows that we have less lower-value flux ($F \sim 0$) and more higher-value flux ($0.5 < F < 0.8$) compared with observations. The flux PDFs from simulation 2048 are consistent with results from simulation 1024, indicating
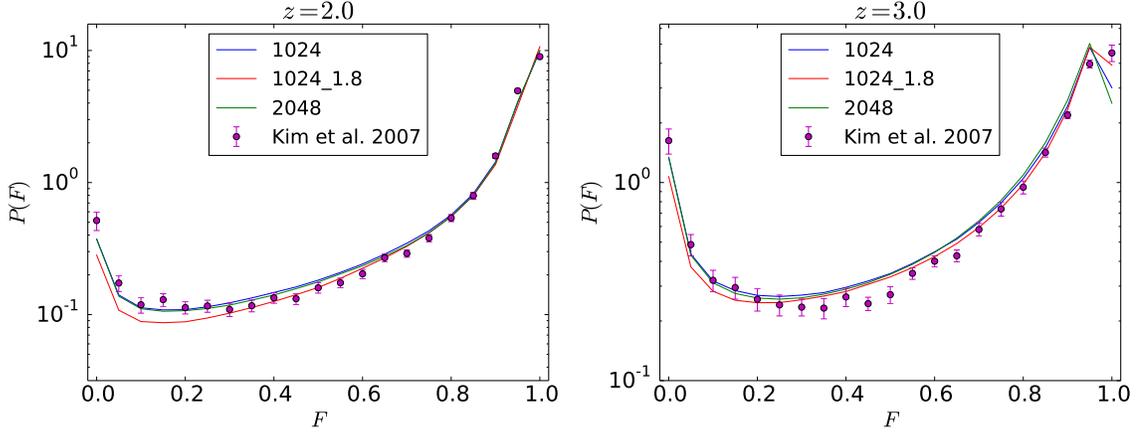
Figure 3.10: H I LAF Flux PDF at redshift $z = 2$ (Left) and $z = 3$ (Right), compared with observations results extrapolated to $z = 2$ and $z = 3$ in Kim et al. (2007). Clear mismatches are visible in both plots.

the convergence of different box sizes.

We also try to rescale the flux F with a power index α, and show PDFs of $F_i^\alpha = \mathrm{e}^{-\alpha \tau_i}$ with a range of α's for simulation 1024 in Fig. 3.11. Taken $z = 2$ for example, with $\alpha = 0.8$ (corresponding to an increase of flux, and a decrease of opacity), the rescaled flux PDF agrees well with data at $0.3 < F < 0.9$, but becomes too low at $F < 0.3$. With $\alpha = 1.1$ (corresponding to a decrease in flux, and an increase of opacity), the rescaled flux PDF agrees well with data at $F < 0.3$, but becomes too high at $0.3 < F < 0.9$. We conclude that the flux PDF from standard LAF simulations has a wrong shape, and that even by rescaling fluxes the mismatch between the simulated and observed flux PDF still exists.
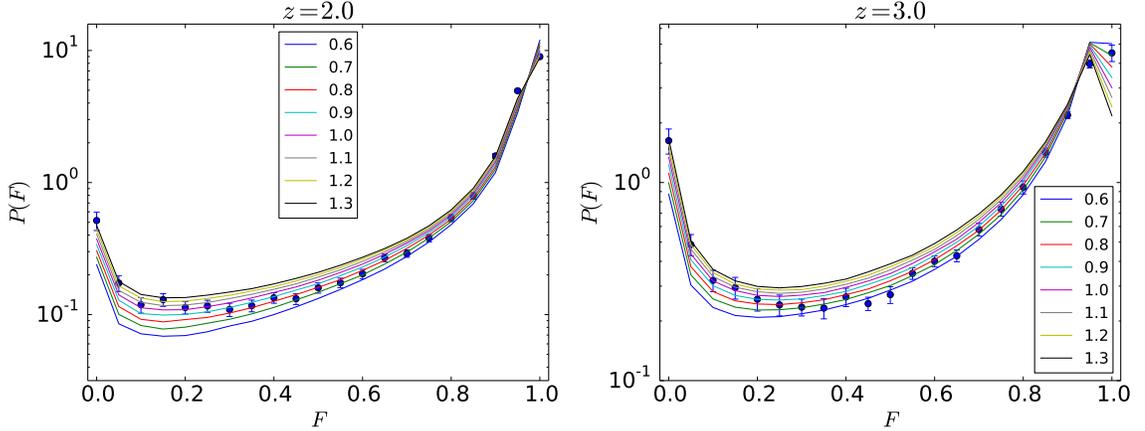
Figure 3.11: Flux PDFs after rescaling by F^α for simulation 1024 at redshift $z = 2$ (Left) and $z = 3$ (Right). The results of different α's are shown. The mismatch still exists.

3.5 Flux power spectrum

Another useful diagnostic of the Lyman alpha forest absorption is the flux power spectrum (FPS). LAF FPS is calculated from spatial correlations of the LAF, and it could be used as an indicator of the density fluctuations at high redshifts (Lukić et al., 2015). We first define the LAF flux contrast as $\delta_F = F/\langle F \rangle - 1$, where $\langle F \rangle$ is the average of flux for all simulated spectra at a certain redshift. Along each line of sight (LOS), we apply Fourier transformation to δ_F in velocity space (Lukić et al., 2015; Tytler et al., 2009):

$$\hat{\delta}(k) = L_v^{-1} \int \delta(v) e^{ikv} dv. \tag{3.3}$$

Here $L_v = LH(z)/(1 + z)$ is the length of each spectrum in velocity space, where L is the simulation box size in comoving Mpc, $H(z)$ is the Hubble constant at redshift z.
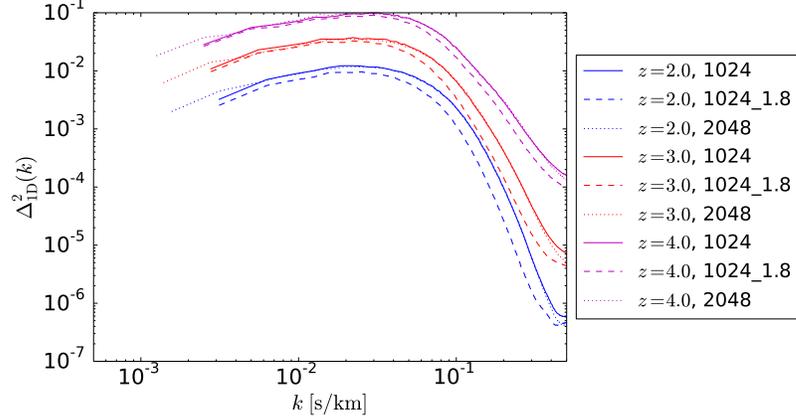
Figure 3.12: The dimensionless power spectrum $\Delta^2_{1D}(k)$ for simulation 1024, 1024_1.8, and 2048 at redshift $z = 2,3,4$.

Then we get the 1D power spectrum $P_{1D}(k)$ in units of s/km:

$$P_{1D}(k) = L_v \left\langle \hat{\delta}(k)\hat{\delta}(k)^* \right\rangle, \tag{3.4}$$

where the average $\langle \rangle$ is over all LOSs at redshift z. And we also calculate the dimensionless power spectrum $\Delta^2_{1D}(k)$ as

$$\Delta^2_{1D}(k) = \frac{k}{\pi} P_{1D}(k). \tag{3.5}$$

The dimensionless power spectrum $\Delta^2_{1D}(k)$ for simulations 1024 and 1024_1.8 at redshift z=2,3,4 are shown in Fig. 3.12. Its lower limit in axis-x is $k_{\text{box}} = 2\pi/L_v$, while its upper limit is the Nyquist frequency $k_{\text{cell}} = k_{\text{box}} \times \frac{N}{2}$, where N is the number of pixels along each LOS (Tytler et al., 2009). It shows that from $z = 4$ to $z = 2$ the power of LAF flux decreases. This is a consequence of the fact that the IGM is becoming more transparent as it expands. At $z = 4$ and $z = 3$, extra He II photoheating in simulation 1024_1.8 results in a higher power in large scale (i.e. lower
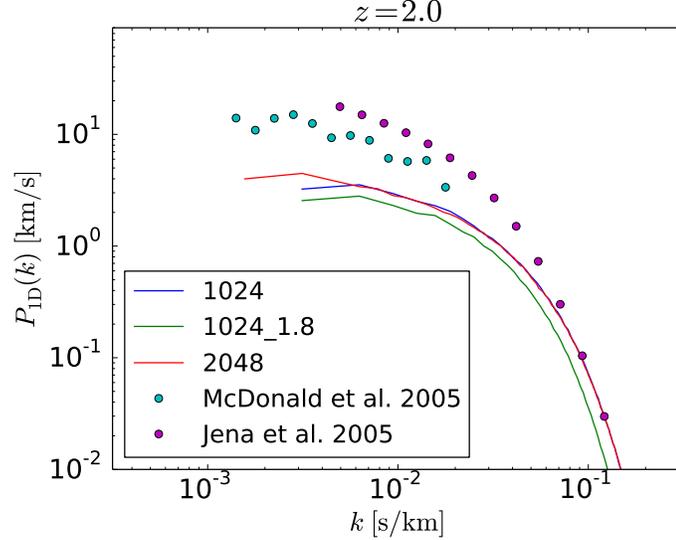
Figure 3.13: 1D power spectrum $P_{1D}(k)$ at redshift $z = 2$. Observations from McDonald et al. (2006) linearly extrapolated to $z = 2$ and from Jena et al. (2005) are also shown.

k), and a smaller power in small scale (i.e. larger k). At $z = 2$, extra photoheating results in a lower power in all scales shown.

In Fig. 3.13 we show the 1D power spectrum $P_{1D}(k)$ for simulation 1024 and 1024_1.8 at redshift $z = 2$. They are compared with observations from McDonald et al. (2006) linearly extrapolated to $z = 2$ and from Jena et al. (2005). It shows that although the power at small scale ($\sim 10^{-1}$ s/km) agrees well with the observations, our simulations, or the standard LAF simulations, miss large scale power. In all the plots showing power spectrum, simulation 2048 has consistent results with simulation 1024, except that 2048 extends to the higher linear scale (i.e. lower wavelength k). It is the discrepancies exhibited in Figs. 3.12 and 3.13 that motivate us to consider the effects of inhomogeneous reionization of He II on the observable properties of the LAF. We examine this in Chapter 5.

Chapter 4

Lyman Alpha Forest simulations with Quasars. I. Thermal and Ionization Evolution

In this chapter, we introduce simulations with both the homogeneous but evolving UV background for H I and He I ionization, and the He II ionizing radiation from quasars, which are treated as time varying point sources. The main parameters of the simulations discussed in this chapter are summarized in Table 4.1. In 1024_80_BG we turn on the UV background from Haardt and Madau (2012) (HM2012) at redshift $z_{\mathrm{BG}} = 15.1$ until the simulation ends ($z = 2.0$). In this simulation, photo-ionization and photo-heating of all 3 primordial species (H I, He I, and He II) is included in the optically thin limit. For the rest 4 simulations, there are three stages in ionizing sources:

a) $99 > z > z_{\mathrm{BG}}$. Here we apply no ionizing background or ionizing sources. Over this redshift range the IGM expands and cools adiabatically. Collapsing structures

(filaments, halos) will heat by adiabatic compression and shocks, and the gas may collisionally ionize.

b) $z_{\mathrm{BG}} > z > z_{\mathrm{QSO}}$. Here we turn on the HM2012 UV background, including photo-ionization and photo-heating for H I, He I, He II, and Compton heating.

c) $z_{\mathrm{QSO}} > z > 2$. Here we add quasars as point ionizing sources into our simulation box, and use MGFLD to transfer the UV radiation from them. The spectrum of quasars in the box starts from 4 Ryd, so they almost only ionize and heat He II. To avoid the duplication of He II ionizing photons, there is no photo-ionization or photo-heating of He II in the UV background at this stage.

Table 4.1: List of simulations in Chapter 4.

Name	Grids	Box size [Mpc/h]	Cell size [kpc/h]	X_{228}	z_{BG}	z_{QSO}	z_{ov}
1024_80_BG	1024^3	80	78	1.0	15.1	N/A	3.15
1024_80_z5	1024^3	80	78	1.0	15.1	5.0	2.57
1024_80_z7	1024^3	80	78	1.0	15.1	6.5	2.53
1024_126_z5	1024^3	126	123	1.0	15.1	5.0	2.77
1024_126_z7	1024^3	126	123	1.0	15.1	6.5	2.75

Notes: X_{228} is the multiplication factor we apply to He II photo-heating rates in HM2012. z_{BG} is the redshift when the UV background is turned on. z_{QSO} is the redshift when we start to place quasars into the simulation box. Photo-ionization and photo-heating of He II in the UV background is turned off when $z < z_{\mathrm{QSO}}$. z_{ov} is the redshift when the volume average of He II fraction $\langle f_{\mathrm{HeII}}\rangle_V$ first drops below 0.01.

4.1 Methodology

In simulations 1024_80_z5, 1024_80_z7, 1024_126_z5, and 1024_126_z7 we place quasars in our simulation box. We assume that the duty cycle of every quasar is 45 Myr, which is consistent with estimates from recent observations (see Table 2 in

Compostella et al. (2013)). Starting from redshift z_{QSO}, we update the positions of quasars in our simulation box every 45 Myr at z_i, where $1 \leq i \leq N$, $z_1 = z_{\mathrm{QSO}}$, z_N is less than 45 Myr to $z = 2$. Then the locations and luminosities of quasars stay constant until z_{i+1}. The procedures to place quasars at redshift $z = z_i$, $1 \leq i \leq N$, are:

1) Find Halos using yt, and load their mass and locations information.

2a) In 1024_80_z7 and 1024_126_z7 when $5 < z \leq 6.5$, we use the quasar luminosity function in Giallongo et al. (2015) (See Eqn. (2) and Table 3). With a fixed bin size $\Delta \log(L_{1450}/(\mathrm{erg/s/Hz})) = 0.25$, we divide the whole range of L_{1450} (corresponding to $M_{1450\_\mathrm{min}} = -18.5$ to $M_{1450\_\mathrm{max}} = -28$) into $N_L = 15$ bins. We then integrate the luminosity function $\frac{d\Phi}{d\log(L_{1450})}(z)[\mathrm{QSOs/Mpc^3/}\log(L_{1450})]$ over bin j and over our simulation volume to get the number of quasars m_j anticipated in luminosity bin j, $1 \leq j \leq N_L$.

2b) For $2 < z_i \leq 5$, for all 4 simulations with quasars we use the quasar luminosity function in Hopkins et al. (2006) (full model, see Eqn. (6)-(10)). With a fixed bin size $\Delta \log(L_{\mathrm{B}}/L_{\odot}) = 0.25$, we divide the whole range of bolometric luminosity ($L_{\mathrm{B\_min}} = 10^{10} L_{\odot}$ to $L_{\mathrm{B\_max}} = 10^{15} L_{\odot}$) into $N_L = 20$ bins. We then integrate the luminosity function $\frac{d\Phi}{d\log(L_{\mathrm{B}})}(z)[\mathrm{QSOs/Mpc^3/}\log(L_{\mathrm{B}})]$ over bin j and over our simulation volume to get the number of quasars m_j anticipated in luminosity bin j, $1 \leq j \leq N_L$.

3) We draw from a Poisson distribution with $\lambda t = m_j$ to get the actual number of quasars we would use n_j, $1 \leq j \leq N_L$.

4a) In 1024_80_z7 and in 1024_126_z7 when $5 < z \leq 6.5$, from $j = N_L$ to $j = 1$, the top n_j heaviest halos without quasars are chosen. For each of the n_j halos,

we draw the specific luminosity L_{1450} of its hosted quasar uniformly from L_{1450} bin j. Its specific luminosity at 1 Ryd L_{ν_0} is deduced from L_{1450} using the following double-power law relation (Vanden Berk et al., 2001; Telfer et al., 2002):

$$L_\nu \propto \begin{cases} \nu^{-0.44} & (\lambda > 1200\text{Å}) \\ \nu^{-1.57} & (\lambda < 1200\text{Å}) \end{cases}.$$

4b) When $2 < z_i \leq 5$, from $j = N_L$ to $j = 1$, the top n_j heaviest halos without quasars are chosen. For each of the n_j halos, we draw the bolometric luminosity of its hosted quasar uniformly from $L_{\rm bol}$ bin j. Its specific luminosity at 1 Ryd L_{ν_0} is calculated by linearly interpolating the $L_{\rm bol}$ - νL_ν table given by the quasar luminosity calculator from Hopkins et al. (2006)[1].

5) The radiation energy of each energy group of each halos is calculated from integration of their spectrum $L_\nu = L_{\nu_0} \left(\frac{\nu}{\nu_0}\right)^{-\alpha}$, where the spectral index α for each halo is drawn from a Gaussian distribution with $\bar{\alpha} = 1.6$ and $\sigma_\alpha = 0.2$. So every quasar has its own spectrum. Here we use 5 energy bins, which are [54.4 eV, 65 eV], [65 eV, 75 eV], [75 eV, 125 eV], [125 eV, 155 eV], [155 eV, 400 eV].

6) Each quasar is put into the simulation at the center of mass of its hosting halo as a constant source for the next 45 Myr. 45 Myr later we remove all the current quasars and go back to step 1).

We run all simulations in this chapter down to $z = 2$.

In Fig. 4.1 we show the observed (Giallongo et al., 2015) and implemented

---

[1]The source code is at `http://www.cfa.harvard.edu/~phopkins/Site/qlf.html`. The results are compiled from about 40 observations. But, as the author says, this code is poorly adapted to handle the rest-frame wavelengths near the Hydrogen absorption edge, ($100 <$ lambda < 912 angstroms, i.e. energies $\sim$1-10 Rydberg), where the cross sections are extremely high and a simple extinction curve application completely obscures that portion of the spectrum. As such, this portion of the spectrum should be treated with care. Hopefully there will be better ways to link a quasar's bolometric luminosity $L_{\rm bol}$ to its specific luminosity L_ν at a specific energy.
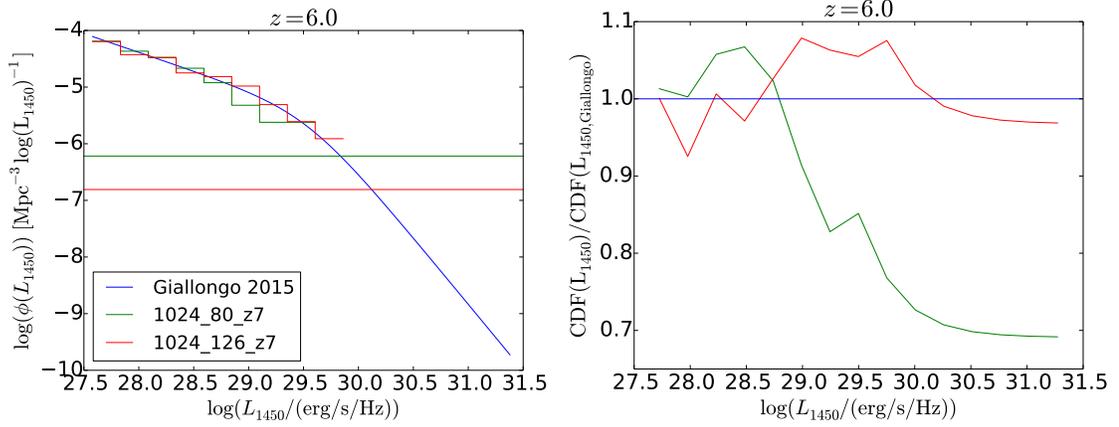
Figure 4.1: Left: Observed vs. implemented specific luminosity function at $z = 6$. The observed luminosity function is from Giallongo et al. (2015). The horizontal lines are the number density corresponding to one quasar in our simulation box, i.e. $1/(80\mathrm{Mpc/h})^3$ for 1024_80_z7 or $1/(126\mathrm{Mpc/h})^3$ for 1024_126_z7. Right: Implemented cumulative distribution function of luminosity divided by observed values.
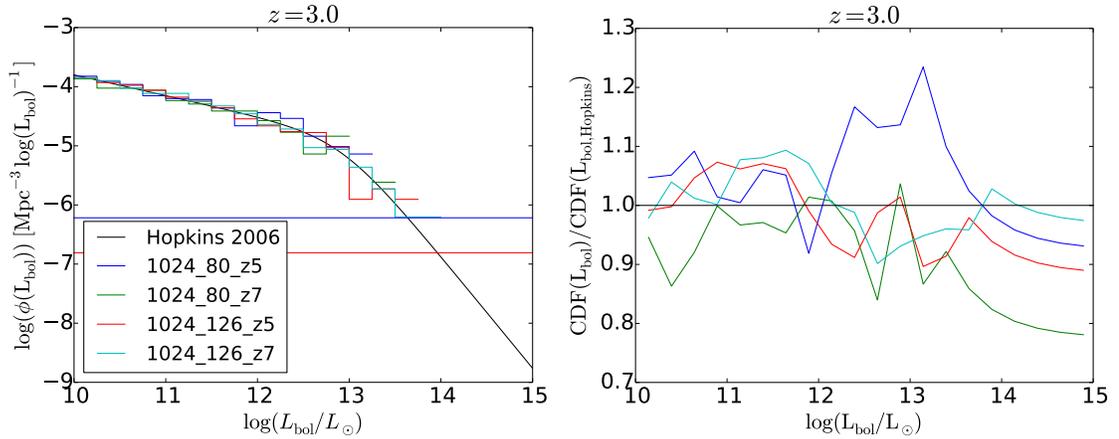
Figure 4.2: Left: Observed vs. implemented bolometric luminosity function at $z = 3$. The observed luminosity function is from Hopkins et al. (2006). The horizontal lines are the number density corresponding to one quasar in our simulation box, i.e. $1/(80\mathrm{Mpc/h})^3$ for 1024_80_z5, 1024_80_z7 or $1/(126\mathrm{Mpc/h})^3$ for 1024_126_z5, 1024_126_z7. Right: Implemented cumulative distribution function of luminosity divided by the observed values.

specific luminosity function for simulation 1024_80_z7 and 1024_126_z7 at $z = 4.0$. The left panel shows that in simulation 1024_80_z7 we have nonzero quasars up to specific luminosity $L_{1450} \sim 10^{29.5}$erg/s/Hz, which approximately corresponds to the number density of 1 quasar per simulation box in our case (represented by the green horizontal line). Since we have a limited number of luminosity bins and we draw from a Poisson distribution in each bin, the implemented number of quasars in every luminosity bin varies around the observed value. The right panel shows the cumulative distribution function of luminosity $\text{CDF}(L_{1450})$ divided by observed values as a function of L_{1450}, where we define $\text{CDF}(L_{1450,i}) = \sum_{j \leq i} L_{1450,j} \times n_j$. It shows that we implement $1 \pm 20\%$ of observed luminosity up to $\sim 10^{29.5} L_{\odot}$ in 1024_80_z7, then the ratio drops since we don't have more luminous quasars in our box due to the limited box size. Note that at other redshifts the ratio could be as high as ~ 3 due to fluctuations in the number of the most luminous quasars. In Fig. 4.2 we show a similar plot at $z = 3$, where we implement the bolometric luminosity function from Hopkins et al. (2006) in all 4 simulations with quasars.

Another interesting quantity related with He II reionization process is the number of ionizing photons. To estimate that, we sum up the radiation energy (in erg/s) in energy group 1 (54.4 to 65 eV) over all current quasars in the box, and divide it by the average energy 59.45 eV (assuming spectral index $\alpha = 1.6$), and multiply it by a time interval (45 Myr for each period with constant sources). In Fig. 4.3 we show the cumulative count of ionizing photons from energy group 1 $N_{\text{ion,HeII,1}}$ divided by the number of He (He I + He II + He III) N_{He}. Jumps in the curve (like 1024_80_z5 at z 3.8) are due to the extra highly luminous quasars drawn from a Poisson distribution.
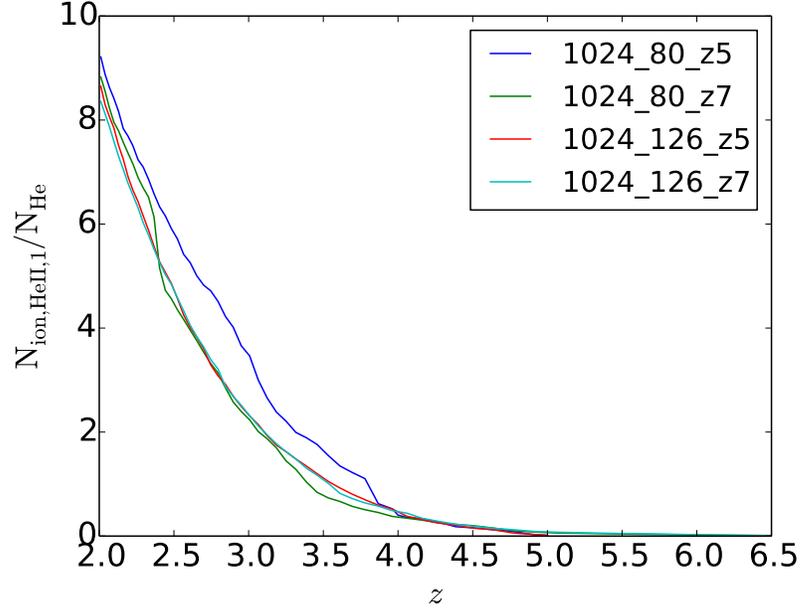
Figure 4.3: Cumulative count of He II ionizing photons from energy bin 1 $N_{ion,HeII,1}$ divided by the number of He (all species) N_{He} versus z.

4.2   Global Evolution Properties

To quantitatively describe the ionization state of the IGM, in each cell we calculate the H I (or He II, He III) density divided by the total density of all H (or He) species, then we take the average value among certain cells to get $\langle f_{HI} \rangle_V$ (or $\langle f_{HeII} \rangle_V$, $\langle f_{HeIII} \rangle_V$). Here $\langle \rangle_V$ represents a volume average over the selected cells. Fig. 4.4 shows the evolution of $\langle f_{HI} \rangle_V$ at various overdensities. Overall the H I fraction drops with time at all overdensities due to Hubble expansion, but cells with higher baryon overdensity Δ have a higher fraction of H I, which is due to the higher recombination rate in denser gas. At each overdensity, there is a redshift interval when $\langle f_{HI} \rangle_V$ is larger in 1024_80_z5 than in 1024_80_BG, and smaller afterwards. That redshift interval corresponds to the time when temperature T is lower in 1024_80_z5 than in 1024_80_BG for that overdensity (see Fig. 4.7 and Fig. 4.8). We expect a higher
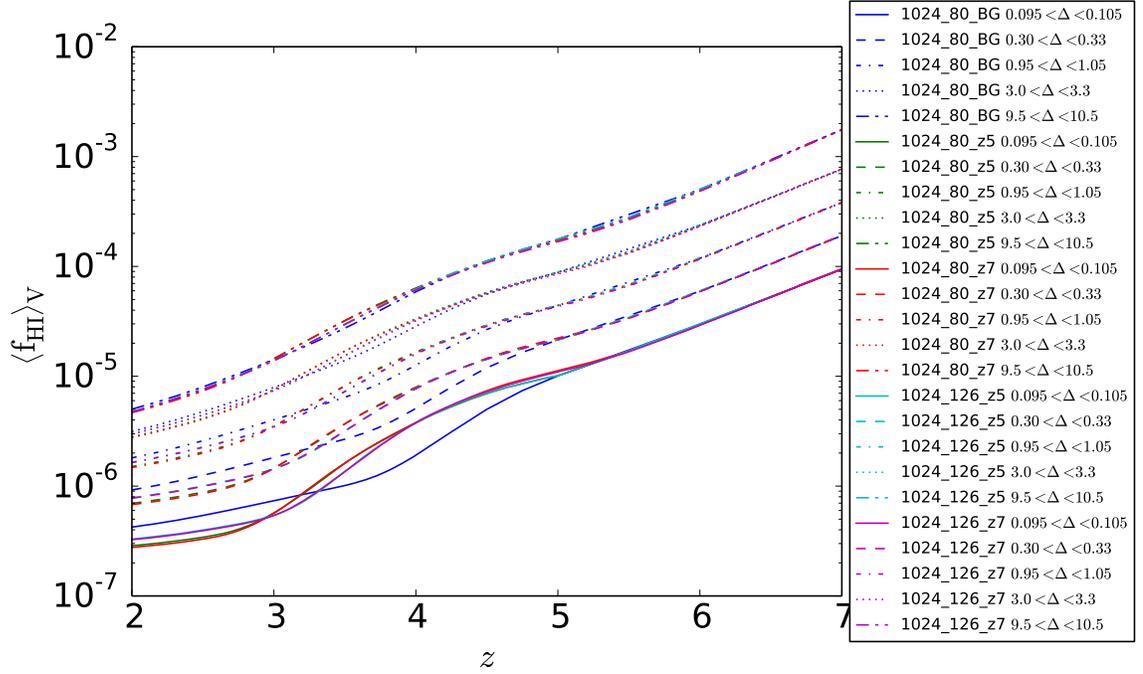
Figure 4.4: Fractions of H I at locations with various overdensities versus redshift. The average is done for cells with baryon overdensity within $1 \pm 5\%$ of 10^{-1}, $10^{-0.5}$, 1, $10^{0.5}$, 10^1, respectively. Values from the same simulation are shown in the same color, while values with the same overdensity are shown in the same line style.

$\langle f_{\mathrm{HI}} \rangle_V$ at lower T, due to the higher recombination rate. Note that H I is ionized by the UVB identically in all simulations. The differences seen in Fig. 4.4 are strictly due to different He II heating histories, which modify the H I recombination rate through its temperature dependence.
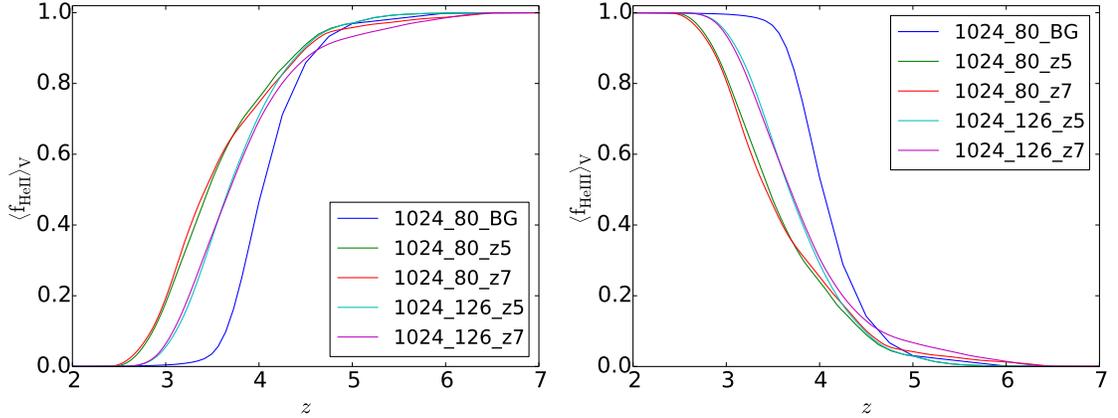
Figure 4.5: Left: Fraction of He II versus redshift. Right: Fraction of He III versus redshift.

Fig. 4.5 shows the evolution of $\langle f_{\mathrm{HeII}} \rangle_V$ and $\langle f_{\mathrm{HeIII}} \rangle_V$ for the whole volume. For simulation 1024_80_BG, there is a small but non-zero He II-ionizing background from $z \sim 15$, but it is not until $z \sim 6$ that the background is strong enough to make a difference. For simulation 1024_80_z5 and 1024_126_z5, the He II reionization does not start until we start to place quasars into the box at $z = 5$. The He II reionization in 1024_80_BG completes at $z \sim 3$, which is earlier than simulations with quasars as point sources. The reason is that in 1024_80_BG the He II photo-ionization and photo-heating background is uniform so all cells start to be ionized right after the UVB is turned on, while in other simulations it takes time for He III bubbles to grow and percolate, and cells are not photo-ionized until there is non-zero radiation energy density (see Fig. 4.9 and Fig. 4.10). He II reionization ends at $z_{\mathrm{ov}} = 2.57$ in

1024_80_z5, and at $z_{\mathrm{ov}} = 2.77$ in 1024_126_z5, where we define z_{ov} to be the redshift when $\langle f_{\mathrm{HeII}} \rangle_V$ first drops below 0.01. We expect that the He II reionization ends earlier in the larger box since it could sample more luminous quasars (see the higher end of implemented luminosity function in Fig. 4.1 Left and Fig. 4.2 Left for example). The larger box simulation gives a more accurate estimate of the reionization history, and $z \sim 2.77$ is consistent with observation results (Dixon and Furlanetto, 2009), which infers the end of He II reionization from He II Lyα opacity measurements.

Another important global property is the temperature of the IGM. Fig. 4.6 shows temperature T vs. overdensity Δ_{b} phase diagrams for 1024_80_z5, 1024_80_BG, and their differences at five redshifts. It shows that at $z > 3$, the temperature distribution in 1024_80_z5 is more scattered than 1024_80_BG, due to the quasars as point ionizing sources. We show the thermal history at mean density of all simulations in Fig. 4.7. Comparing with 1024_80_BG, 1024_80_z5 peaks later at $z \sim 2.7$ and has a higher peak temperature. The IGM temperature is directly related with the photo-heating rate, and the He II photo-heating rate is proportional to He II density and radiation energy density. So the IGM temperature is related with the reionization history. Due to the extra time it takes for He III bubbles to expand and percolate, at $z \sim 2.7$ $\langle f_{\mathrm{HeII}} \rangle_V \sim 10\%$ in 1024_80_z5 while much less He II is in 1024_80_BG (Fig. 4.5). So we expect that the photo-heating rate in 1024_80_z5 is higher than the optically thin value in 1024_80_BG, considering that both simulations draw from the same luminosity function. We also show temperature from simulation 1152, which is a H I reionization simulation with galaxies as point sources described in Chapter 7.

Comparing with 1024_80_z5, 1024_126_z5 peaks earlier at $z \sim 3$ and has a lower peak temperature. It is earlier because in the larger box there are more luminous
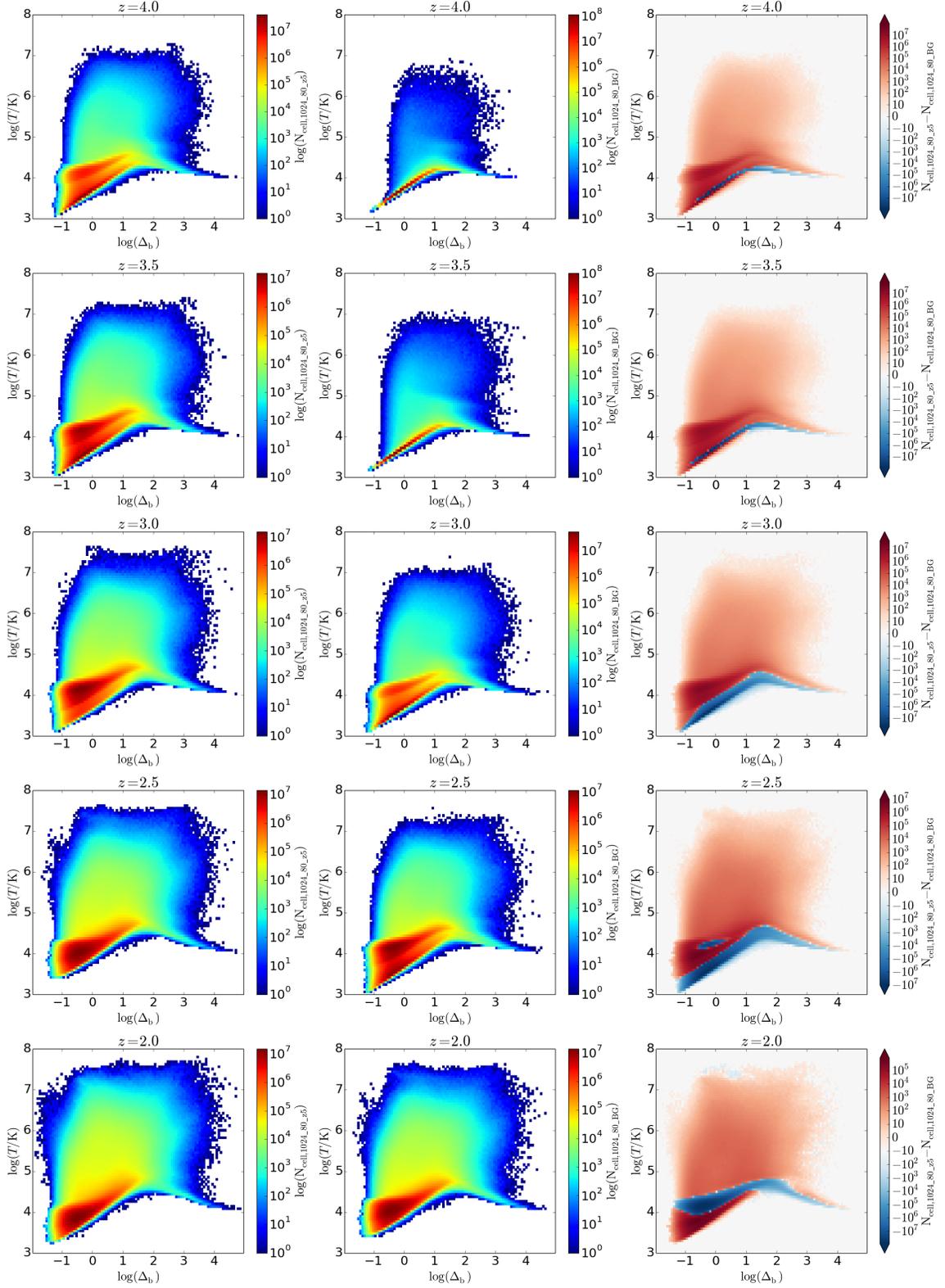
Figure 4.6: Temperature vs. overdensity phase diagrams for 1024_80_z5 (Left) and 1024_80_BG (Middle), and their differences (Right). From top to bottom, they are at $z = 4.0$, 3.5, 3.0, 2.5, and 2.0, respectively.
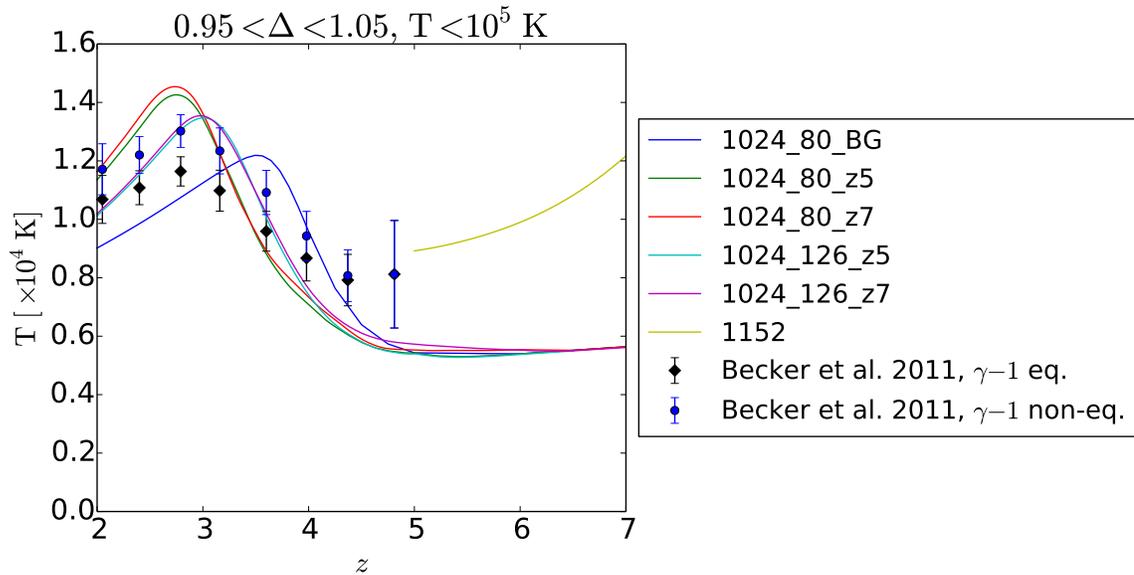
Figure 4.7: Temperature evolution at the mean density. Note that values shown here are the average temperature of cells satisfying $0.95 < \Delta < 1.05$ and $T < 10^5$K. Observational values from Becker et al. (2011) are also shown. Simulation 1152 is a H I reionization simulation and is described in detail in Chapter 7. Note that in this plot we are comparing apples and oranges. The solid lines are taken directly from the simulations. The data points involve an analysis of H I Lyman α forest spectra, which is an indirect and inexact measure of the IGM temperature.
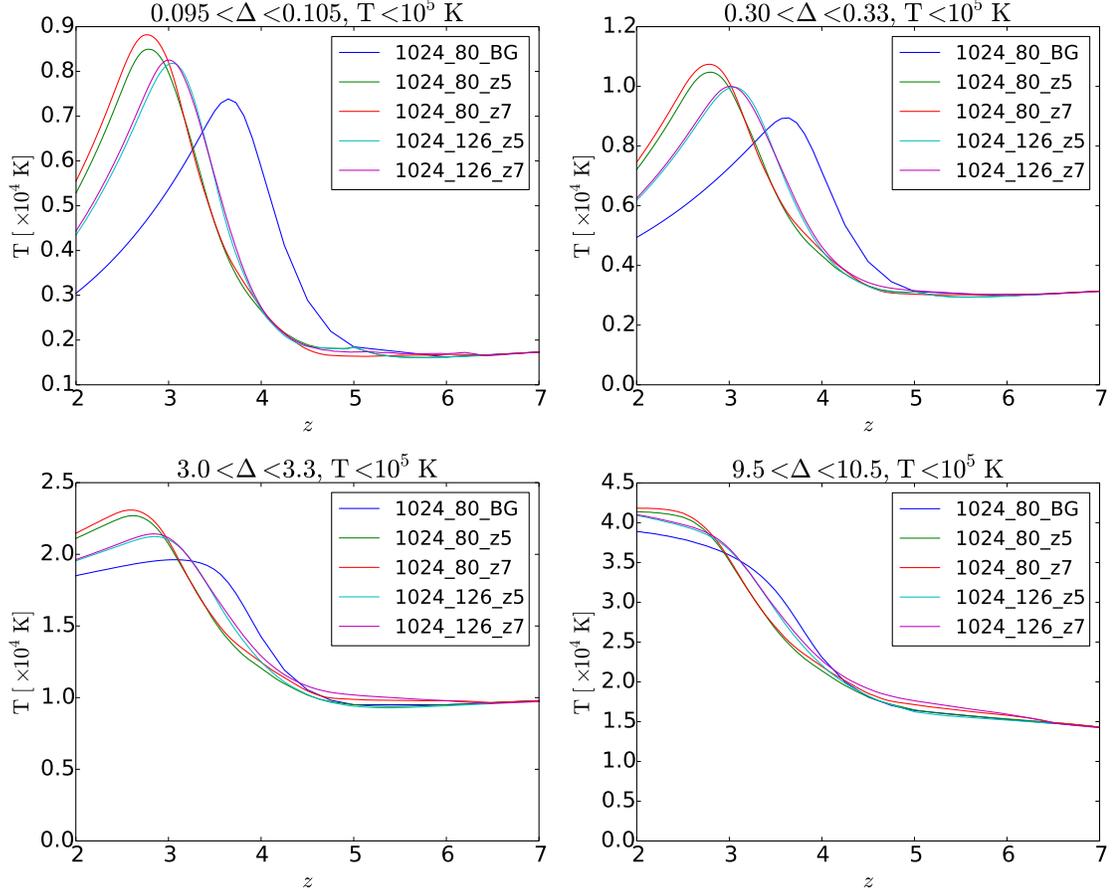
Figure 4.8: Temperature evolution at various densities. From left to right, top to bottom, we show the average temperature of cells with baryon over-density within $1 \pm 5\%$ of 10^{-1}, $10^{-0.5}$, $10^{0.5}$, 10^{1}. Here we only consider cells with $T < 10^5$K.

quasars. Becker et al. (2011) measure the curvature of 61 high-resolution QSO spectra, and convert it into temperature after calibrating with a set of hydrodynamic simulations. Results labeled as "$\gamma - 1$ eq." in the plot are calibrated using simulations with the assumption of ionization equilibrium, while results labeled as "$\gamma - 1$ non-eq." are calibrated with non-equilibrium simulations. Although their results are close to ours in magnitude, they are arrived at by different methods. Our results are taken directly from simulations, instead of following their procedures and calculating curvatures of spectra. A more direct way to compare observed and simulated thermal history is to compute b_σ from observed and artificial spectra respectively, which we will show in Section 5.6.

Temperature evolutions at various overdensities are shown at Fig. 4.8. Overall a larger overdensity corresponds to a higher temperature at all redshifts, which is consistent with the equation of state of the IGM (Eqn. 3.1). Another rough trend is that the time of peak temperature for each simulation delays at larger baryon overdensities. This is due to the fact that high density regions have higher recombination rates and take more time to ionize.

To get a better view of spatial variations of physical quantities, we make several slices. In Fig. 4.9 we show slices of He II fraction f_{HeII}, He III fraction f_{HeIII}, temperature T, He II photo-ionization rate Γ_{HeII}, and He II photo-heating rate $\mathscr{H}_{\text{HeII}}$ at 4 redshifts ($z = 4.0, 3.5, 3.0, 2.0$) for simulation 1024_80_z5. The evolution of f_{HeII} and f_{HeIII} shows the process of He III bubble expanding and percolation. Notably there is a large halo in the center of the volume, and there is always a quasar there according to our scheme, but its emissivity changes every 45 Myr. At redshift $z \sim 3$ most of the He II has been ionized. At redshift $z \sim 2$ the He II fraction drops below

10^{-3} in most regions, while most remaining He II resides in filaments. From those slices our periodic boundary conditions are evident. The slice of temperature at $z = 3$ shows clearly an He III bubble in the center and that the ionizing front has a high temperature. The temperature directly behind the ionization front is higher than in the center of the bubble. This is a consequence of the fact that the IGM is optically thick to He II ionizing photons at the I-front, and absorbs virtually all the radiation energy flux, as opposed to only a fraction of it in the optically thin interior (Abel and Haehnelt, 1999).

Overall, knots are hotter than filaments, and filaments are hotter than voids. The He II photo-ionization rate Γ_{HeII} includes contribution from all 5 He II ionizing radiation energy groups, and can be viewed as a weighted sum of 5 radiation energy densities (Fig. 4.10). Regions with high Γ_{HeII} expands as more and more quasars contribute their ionizing photons to the IGM. The last row shows the He II photo-heating rate $\mathscr{H}_{\mathrm{HeII}}$. Note that in 1024_80_z5 photons from quasars only contribute to photo-heating of He II, but not H I or He I. As $\mathscr{H}_{\mathrm{HeII}}$ is proportional to both He II density and radiation energy density, regions with high values of $\mathscr{H}_{\mathrm{HeII}}$ are those at the outside edge of He III bubbles, as shown in plots.

In Fig. 4.10 we show slices of radiation energy densities of all 5 energy groups of 1024_80_z5 at several redshifts. It shows that at the same redshift (i.e. same column), the radiation energy density in the higher energy bin expands to a larger area, which is due to the lower He II cross-section as a function of photon energy. As time increases from left to right panels, old quasars are replaced by new quasars and they all contribute to the radiation energy in the IGM.

To compare the effect of He II ionizing background in 1024_80_BG versus point
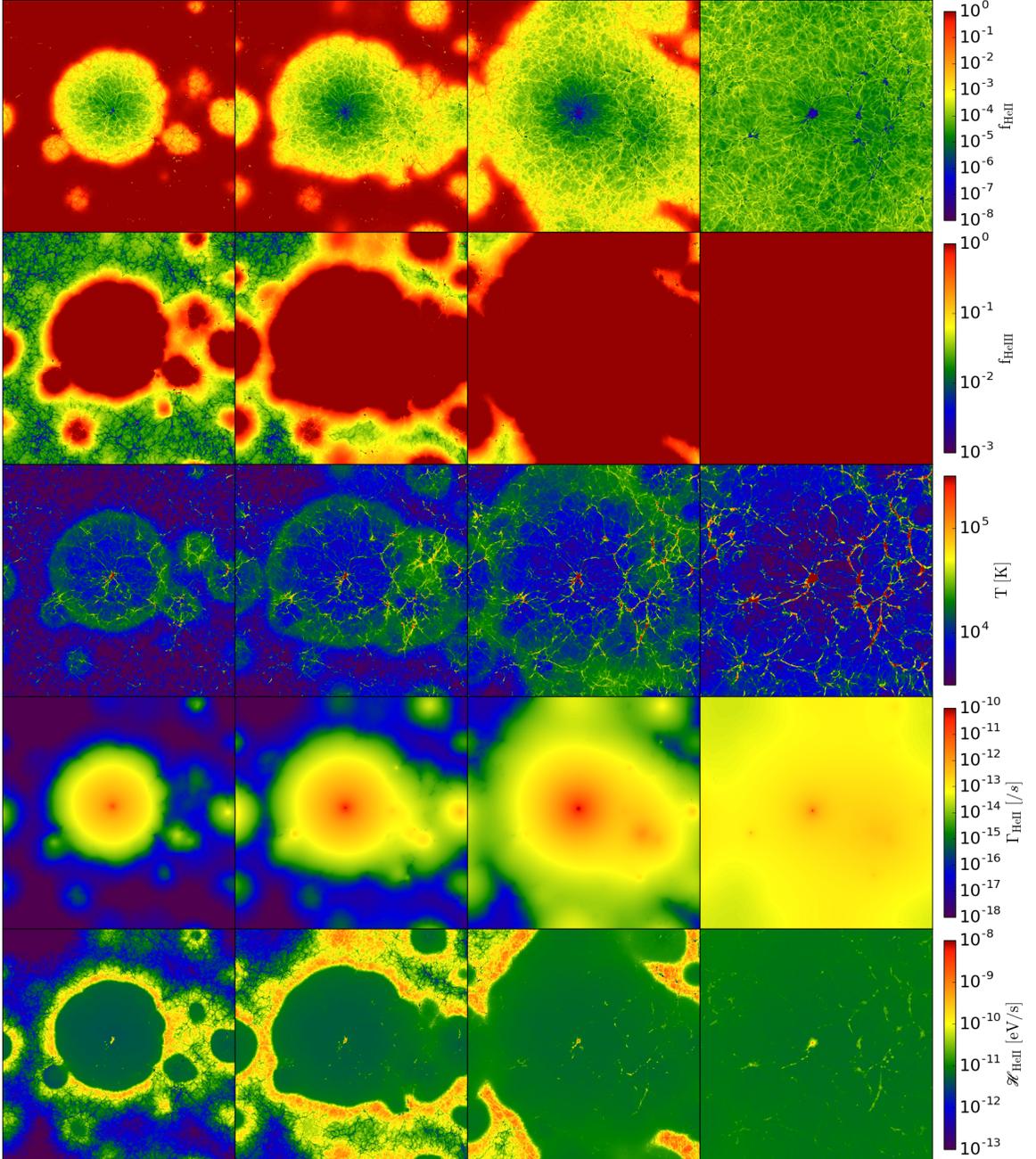
Figure 4.9: Slices of 5 quantities as a function of redshift in 1024_80_z5. From top to bottom we show He II fraction f_{HeII}, He III fraction f_{HeIII}, temperature T, He II photo-ionization rate Γ_{HeII}, and He II photo-heating rate $\mathscr{H}_{\mathrm{HeII}}$. From left to right they are at $z = 4.0$, $z = 3.5$, $z = 3.0$, and $z = 2.0$, respectively. All slices are made at axis z=0.5 plane.
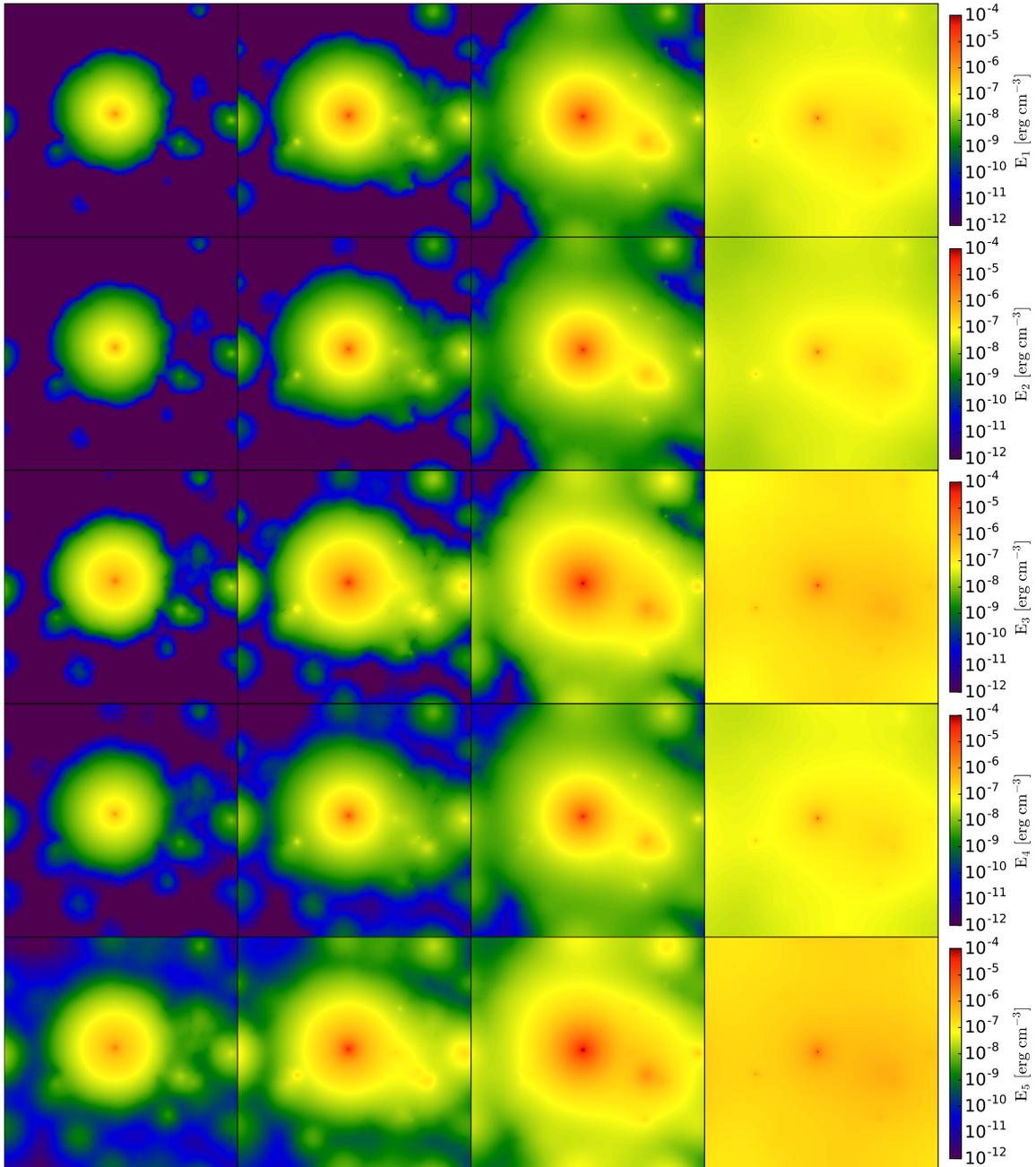
Figure 4.10: Slices of radiation energy densities of 1024_80_z5. From top to bottom we show 5 radiation energy densities: E_1 [54.4 eV, 65 eV], E_2 [65 eV, 75 eV], E_3 [75 eV, 125 eV], E_4 [125 eV, 155 eV], and E_5 [155 eV, 400 eV]. From left to right they are at $z = 4.0$, $z = 3.5$, $z = 3.0$, and $z = 2.0$, respectively. All slices are made at axis z=0.5 plane.
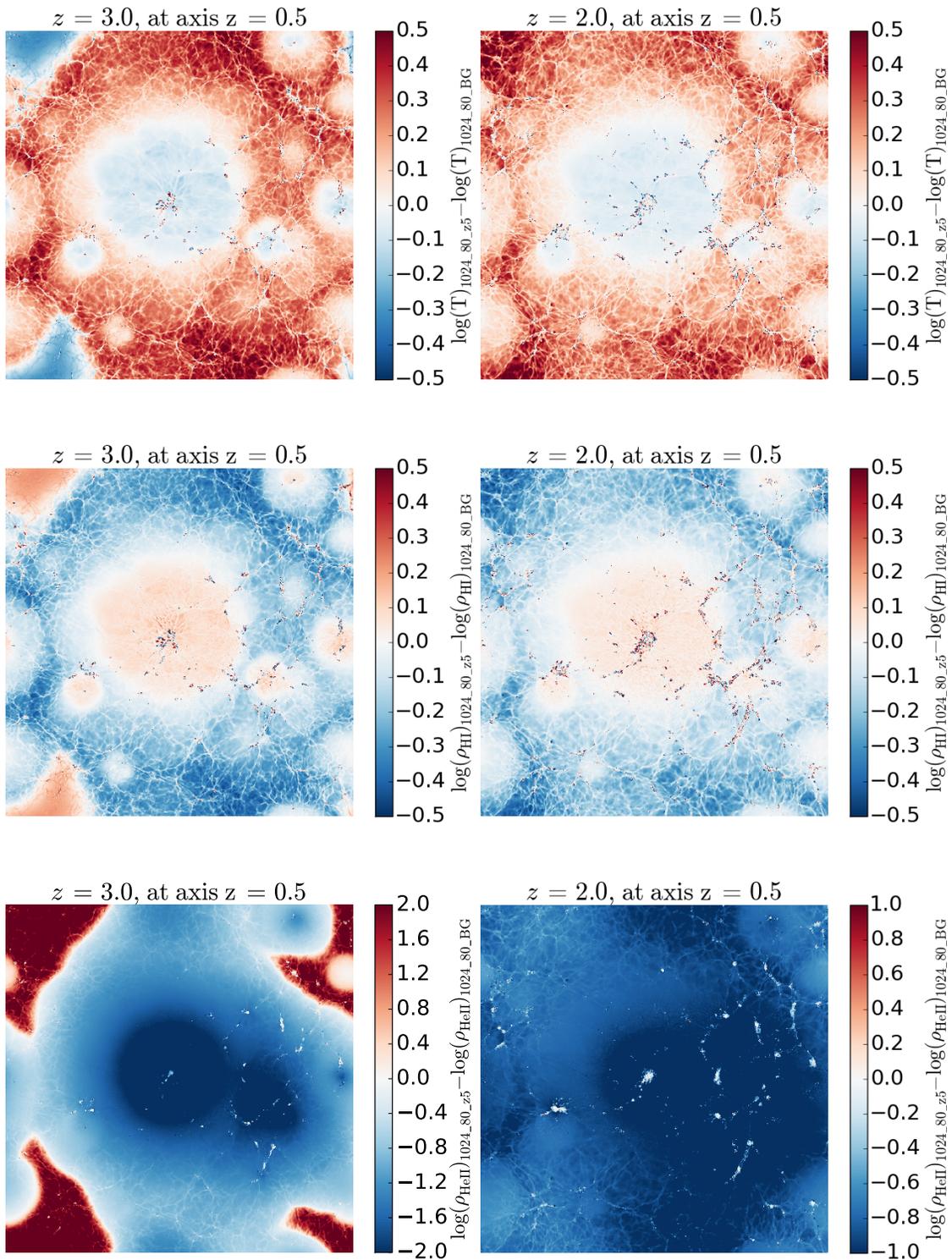
Figure 4.11: Difference of slices of temperature (Top, in K), H I density (Middle, in g/cm³), and He II density (Bottom, in g/cm³) in log scale between 1024_80_z5 and 1024_80_BG at $z = 3$ (Left) and $z = 2$ (Right). Note that 0 value corresponds to white color in the plot.

source quasars in 1024_80_z5, we make several slices in both simulations and calculate the ratio of them (i.e., difference in the log). In Fig. 4.11 we show the difference of log(temperature) (Top), log(ρ_{HI}) (Middle), and log(ρ_{HeII}) (Bottom) between 1024_80_z5 and 1024_80_BG at axis z=0.5 plane at redshift $z = 3$ (Left) and $z = 2$ (Right). Positive differences are shown as red, while negative differences are shown as blue. At redshift $z = 3$, the regions at ionization fronts in 1024_80_z5 are hotter than 1024_80_BG. Temperature is lower in the corners, where little He II ionizing photons have reached yet. At $z = 2$ those regions are at the ionization front and are hotter than 1024_80_BG. Temperature is a bit lower within He III bubbles, which is because He II is highly ionized there and the photo-heating rate there is lower than the average optically thin value. The H I density plot is like the inverse plot of temperature, due to the inverse relation between temperature and H I recombination rate. At $z = 3$, He II density is much lower in He III bubbles in 1024_80_z5 and much higher outside the bubbles comparing with 1024_80_BG. In other words, fluctuations in He II (or He III) density field during the He II reionization are better captured by 1024_80_z5 where quasars are point sources. At $z = 2$, most regions in 1024_80_z5 have lower He II density than 1024_80_BG, except the knots in the cosmic web, where the temperature is so high that both simulations have very low density of He II (Fig. 4.9).

4.3 Effect of high z quasars

Giallongo et al. (2015) find 22 AGN candidates and derive the UV luminosity function in the redshift interval $4 < z < 6.5$. We use their results in simulations 1024_80_z7 and 1024_126_z7 from $z = 6.5$ to $z = 5$, in addition to quasars at $z < 5$ from the luminosity function in Hopkins et al. (2006).

In Fig. 4.5, we see that the earlier placement of quasars in 1024_80_z7 and 1024_126_z7 results in an earlier start of He II reionization, and the faction of He III is slightly higher in 1024_80_z7 than 1024_80_z5 until $z \sim 3.7$, then it's reversed. This reverse is due to the jump in the number of highly luminous quasars in 1024_80_z5 (Fig. 4.3). But there are similar jumps in 1024_80_z7 later on so it doesn't affect the end of He II reionization very much. Fig. 4.7 shows that 1024_80_z7 and 1024_126_z7 have an earlier increase in the IGM temperature, but their thermal histories are very close to 1024_80_z5 and 1024_126_z5 respectively. For cells with mean density, the largest difference in the temperature is ~ 200K (Fig. 4.7). Similar results could be seen in other overdensities (Fig. 4.8). We conclude that high redshift quasars ($5 < z < 6.5$), although they make a difference in the beginning of the reionization and thermal history, do not affect the later He II reionization process very much.

We have seen how quasars modeled as point sources affect the thermal and ionizing history in a different way than a homogeneous UV background. In the next two chapters we will generate artificial spectra and examine how quasars affect H I and He II Lyα forest.

Chapter 5

Lyman Alpha Forest simulations with Quasars. II. Spectroscopic Analysis of H I Lyα Forest

In this chapter we explore the observational signatures of inhomogeneous He II reionization on the hydrogen Lyman α forest. To this end we construct synthetic absorption spectra from the simulations presented in chapter 4 and subject them to a variety of analyses discussed below. Specifically, for simulations 1024_80_BG, 1024_80_z5, and 1024_80_z7, we generate 1024^2 light rays parallel with z-axis from axis z=0 plane to axis z=1 plane. We then generate synthetic H I and He II Lyα absorption spectra for each light ray using yt, and analyze them. The scripts we use for this and all subsequent analyses are presented in the Software appendix. In this chapter we confine our discussion to H I Lyα forest. We present He II Lyα results in Chap. 6.

5.1   Synthetic H I absorption spectra
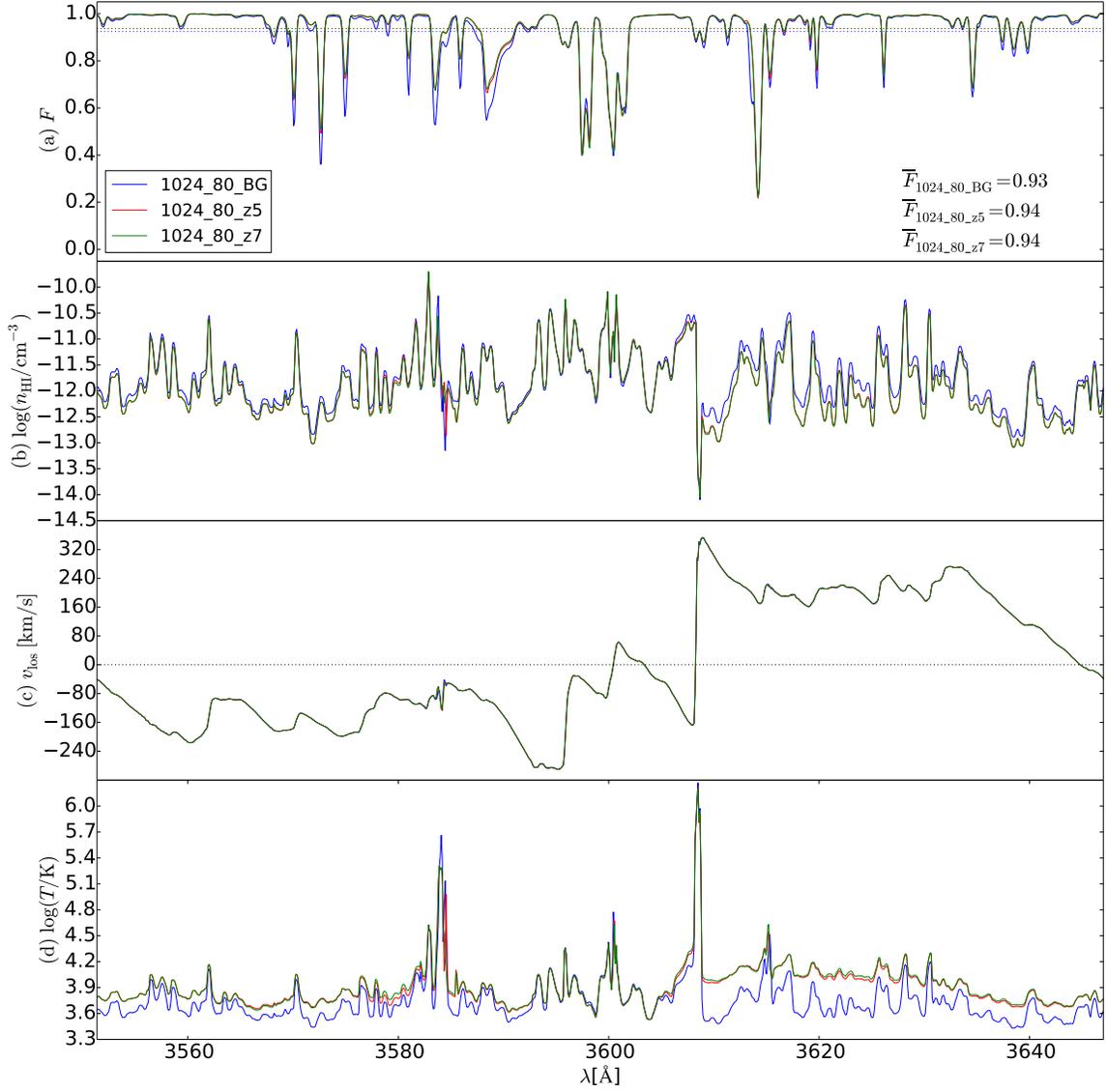
Figure 5.1: Quantities along a line of sight from $z = 2.00$ to $z = 1.92$. The same line of sight is used for all simulations shown. (a) Normalized flux. The average flux of this spectrum is shown at the lower right corner. (b) H I number density. (c) Peculiar velocity along line of sight. The black dotted horizontal line locates at $v_{\mathrm{los}} = 0$. (d) Temperature. Same color is used for one simulation across panels. In this line of sight 1024_80_z5 has higher temperature, less H I, and less absorption than 1024_80_BG.

Fig. 5.1 shows one light ray from $z = 2.00$ to $z = 1.92$ (i.e. 80 Mpc/h comoving,

the size of our simulation box) and its normalized H I Lyα flux F, H I number density n_{HI}, peculiar velocity along line of sight v_{los}, and temperature T. Each light ray has 1024 pixels, which is equal to our number of grids in one dimension. To see the effect of distributed sources, the same line of sight is shown for all simulations for comparison. It shows that typically 1024_80_z5 has higher temperature, less H I, and less absorption (i.e. larger normalized flux) comparing with 1024_80_BG. The peculiar velocities for both simulations are almost the same at each pixel. Fig. 5.2 shows another light ray from $z = 4.00$ to $z = 3.84$. A similar plot from $z = 4$ to $z = 2$ is generated with 22 data dumps (Fig. 5.3). It shows that at $z > 2.5$, in this line of sight 1024_80_z5 has lower temperature, more H I, and more absorption than 1024_80_BG, but they become close later on. The reason for this difference in temperature evolution has been discussed in Chapter 4.
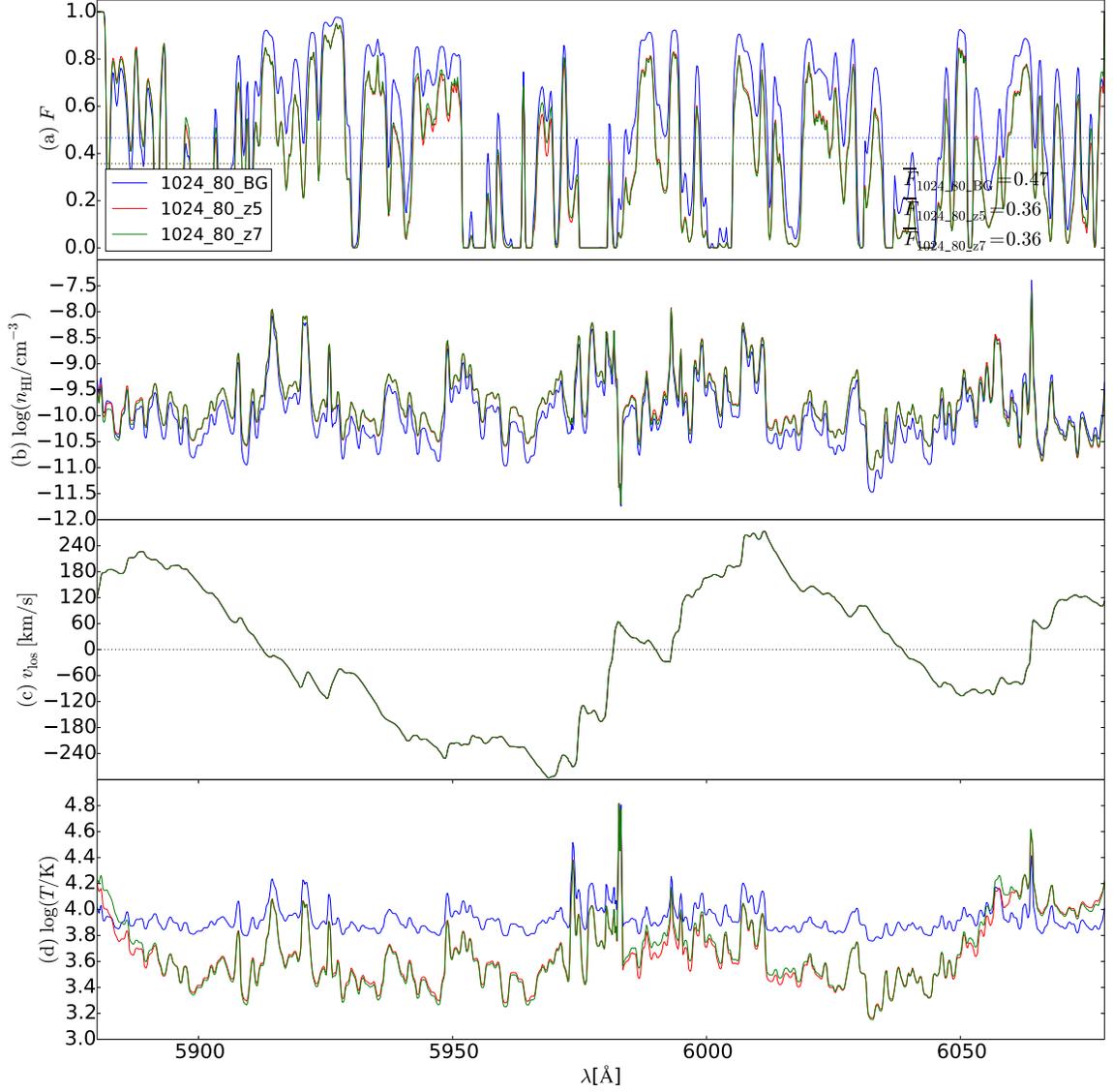
Figure 5.2: Quantities along a line of sight from $z = 4$ to $z = 3.84$. The same line of sight is used for all simulations shown. (a) Normalized flux. The average flux of this spectrum is shown at the lower right corner. (b) H I number density. (c) Peculiar velocity along line of sight. The black dotted horizontal line locates at $v_{\rm los} = 0$. (d) Temperature. Same color is used for one simulation across panels. In this line of sight 1024_80_z5 has lower temperature, more H I, and more absorption than 1024_80_BG.
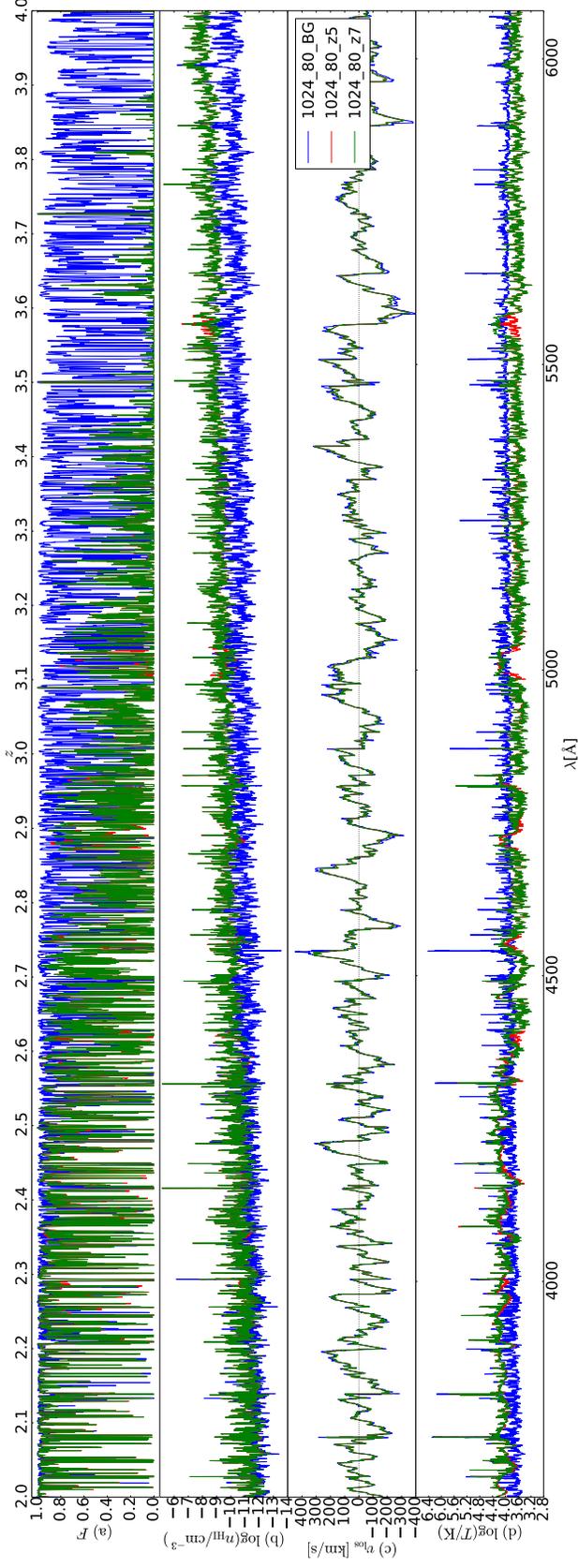
Figure 5.3: Quantities along a line of sight from $z = 4$ to $z = 2$. The same line of sight is used for all simulations shown. (a) Normalized flux of H I LAF. (b) H I number density. (c) Peculiar velocity along line of sight. The black dotted horizontal line locates at $v_{los} = 0$. (d) Temperature. Same color is used for one simulation across panels.

5.2 Optical depth
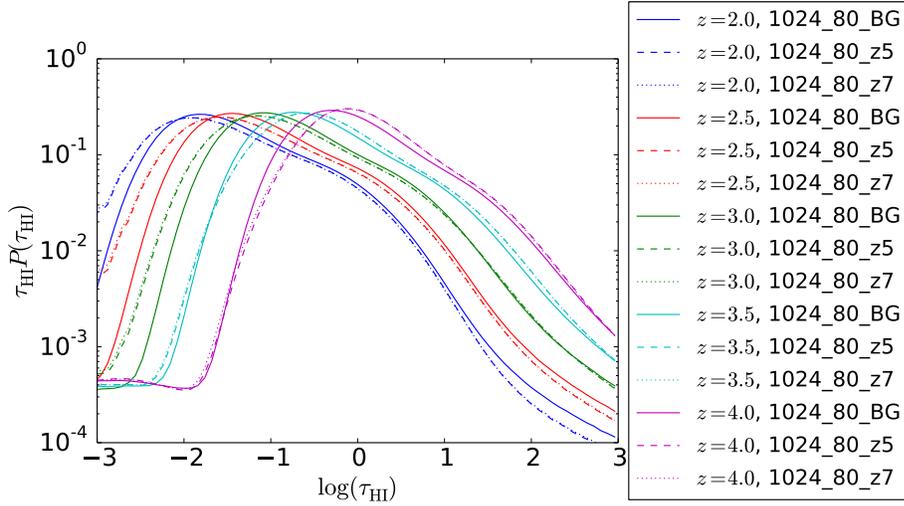
Figure 5.4: The distribution of H I opacity τ_{HI} at different redshifts. Lines with the same line style are from the same simulation, while lines with the same color are at the same redshift.

The normalized flux F can be written as $F_i = e^{-\tau_i}$, where τ_i is the optical depth for pixel i. The distribution of H I opacity τ_{HI} at different redshifts is shown in Fig. 5.4. It shows that the location of the peak decreases over time, as H I is being ionized and the IGM becomes more transparent. At $z > 3.5$ 1024_80_z5 has more higher-value τ_{HI} and less lower-value τ_{HI} than 1024_80_BG, which is due to the lower IGM temperature in 1024_80_z5. At $z < 3.5$, it's reversed, since the IGM temperature in 1024_80_z5 is higher (see Fig. 4.7). Other than the slight shift in the peak of the optical depth distribution function due to temperature, the two kinds of simulations give very similar curves. Our results for 1024_80_BG are consistent with results in Zhang et al. (1997).

We also calculate the effective H I opacity $\tau_{\mathrm{HI}}^{\mathrm{eff}}$, where τ^{eff} is defined as $e^{-\tau^{\mathrm{eff}}} = \langle F \rangle$. For each of 1024^2 line of sights across the simulation box at a certain redshift,
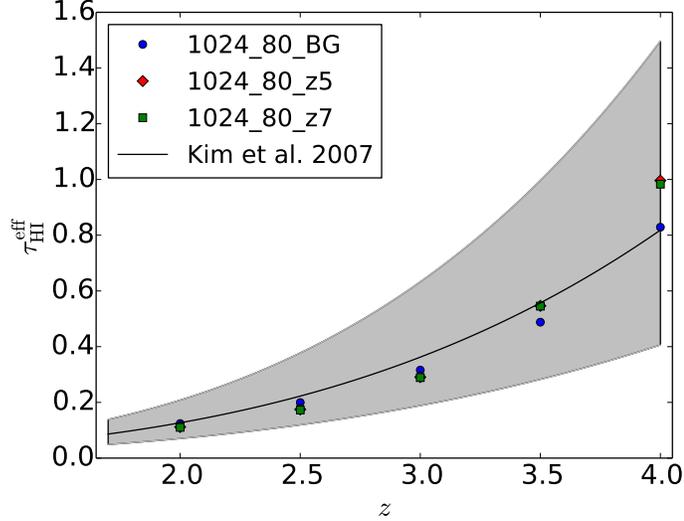
Figure 5.5: Effective H I opacity $\tau_{\mathrm{HI}}^{\mathrm{eff}}$ as a function of z. For each line of sight across the simulation box, we calculate its $\tau_{\mathrm{HI}}^{\mathrm{eff}}$ from its mean flux. The average value over 1024^2 line of sights at certain redshifts are shown. We also show the best fit of $\tau_{\mathrm{HI}}^{\mathrm{eff}}(z)$ from Kim et al. (2007). Their error bars are calculated using a variation of the bootstrap method (See Appendix B of McDonald et al. (2000) for details).

we calculate its τ^{eff} from its mean flux. At 5 redshifts we do such analysis and the average values are shown in Fig. 5.5. The best fit of $\tau_{\mathrm{HI}}^{\mathrm{eff}}(z)$ from Kim et al. (2007) is also shown. All simulation values are well within the 1 σ observational band, and in fact quite close to the mean values. At $z > 3.5$ $\tau_{\mathrm{HI}}^{\mathrm{eff}}$ is larger in 1024_80_z5 than in 1024_80_BG, while it's reversed at lower redshifts. The absolute difference between them reduces with time. This again is due to the relative change of IGM temperature.

To get an idea of line-of-sight variations in $\tau_{\mathrm{HI}}^{\mathrm{eff}}$, we show the distribution of $\tau_{\mathrm{HI}}^{\mathrm{eff}}$ in the 1024^2 sightlines at 5 redshifts in Fig. 5.6. It shows a similar trend as the distribution of τ_{HI} (Fig. 5.4) across redshifts, where $z \sim 3.5$ is the time when the reverse happens. The largest differences between the 3 simulations of the distribution functions occurs at $z = 4$, and diminishes with decreasing redshift so that by $z = 3$,
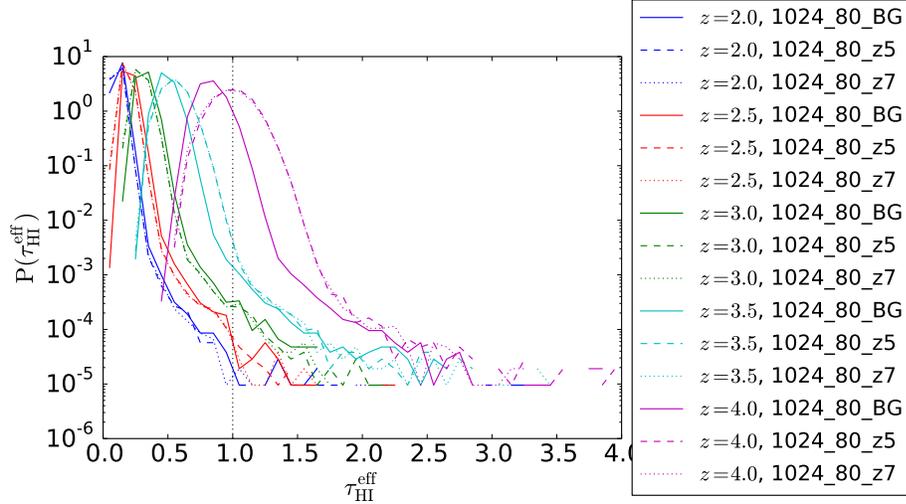
Figure 5.6: PDF of effective H I opacity $\tau_{\mathrm{HI}}^{\mathrm{eff}}$ at redshift $z = 4$, 3.5, 3, 2.5, and 2. The vertical dotted line is at $\tau_{\mathrm{HI}}^{\mathrm{eff}} = 1$. Lines with the same line style are from the same simulation, while lines with the same color are at the same redshift.

they are very similar. This is because the temperature differences between the "BG" and "z" simulations are the largest at $z = 4$. The patchy temperature differences which create patchy H I density differences (Fig. 4.11) seem to have little effect on the shapes of the distributions functions. According to the convention to define $\tau^{eff} < 1$ as optically thin, all simulations are optically thin to H I ionizing photons in the vast majority of sightlines for $z \leq 3.5$, which means a UV background is a good approximation in this period.

5.3   Flux PDF

With all the normalized flux F_i's, we have generated the flux probability distribution function (PDF) $P(F)$. We also compare our results with the observations in Kim et al. (2007) extrapolated to $z = 2$ (Fig. 5.7), $z = 2.5$ and $z = 3.0$ (Fig. 5.8).
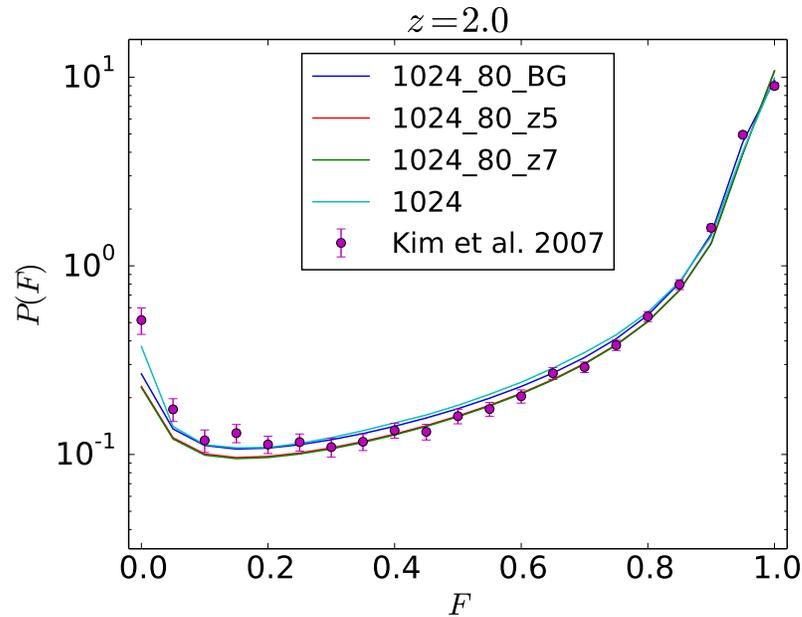
Figure 5.7: H I Lyα flux PDF at $z = 2$. Simulation 1024 is a 1024^3, 20 Mpc/h, UV background-only simulation that is discussed in Chapter 3. Data points are linearly interpolated from Kim et al. (2007).
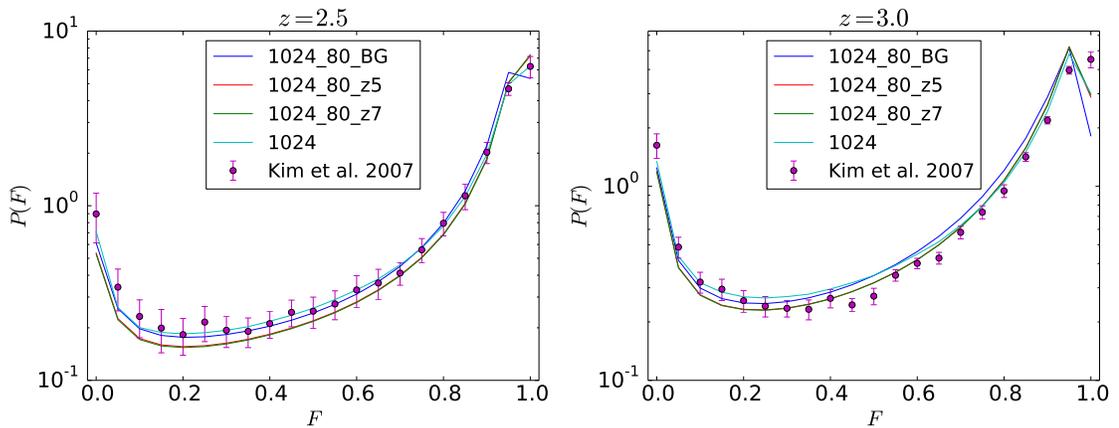
Figure 5.8: H I Lyα flux PDF at $z = 2.5$ (Left) and $z = 3.0$ (Right). Simulation 1024 is a 1024^3, 20 Mpc/h, UV background-only simulation that is discussed in Chapter 3. Data points are linearly interpolated from Kim et al. (2007).

Calculations of $P(F)$ follow the convention in Kim et al. (2007), where we divide flux of all pixels into bins with width $\Delta F = 0.05$. Pixels with $F < 0.025$ or $F > 0.975$ are counted in the $F = 0$ bin and the $F = 1.0$ bin, respectively. Note that all our results shown here are directly from synthetic spectra, without normalizations to observed mean fluxes or opacities. At $z = 2$ the figure shows that the shapes of the simulated PDFs are qualitatively similar to the observations, and they are in better agreement with data than the small-box simulations (Fig. 3.10). But there is still some discrepancy. Specifically, we have less zero-value flux compared with observations. We also show result from simulation 1024, which is a 1024^3, 20 Mpc/h, UV background-only simulation that is discussed in Chapter 3. Its results basically agree with 1024_80_BG, except that it has more pixels at $F \sim 0$. So a high resolution is necessary to produce high density absorbers so that we have more flux with $F \sim 0$. At $z = 2.5$ all simulations agrees with data very well, except the lower end. At $z = 3.0$ the discrepancy is larger, but 1024_80_z5 is in better agreement with the observational results than 1024_80_BG.

The $P(F)$ of all 3 simulations at 5 redshifts are shown in Fig. 5.9. We show the ratios of $P(F)$ between 1024_80_z5 and 1024_80_BG at 5 redshifts in Fig. 5.10. It shows that 1024_80_z5 has more lower-value F, and less high-value F at $z > 3.5$. The situation reverses at $z < 3.5$, when the higher IGM temperature in 1024_80_z5 causes less H I, less absorption, and thus more transmission.
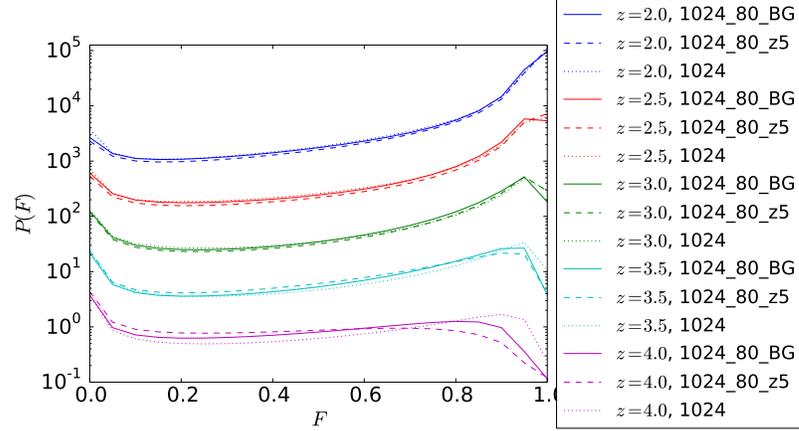
Figure 5.9: PDF of H I Lyα flux $P(F)$ at 5 redshifts. Lines with the same line style are from the same simulation, while lines with the same color are at the same redshift. Simulation 1024 is a 1024^3, 20 Mpc/h, UV background-only simulation that is discussed in Chapter 3. Results of 1024_80_z7 are not shown here since they are very close to results of 1024_80_z5. From top to bottom, we show $P(F)$ at $z = 2$, 2.5, 3, 3.5, 4, and they are multiplied by a factor of 10^4, 10^3, 10^2, 10^1, 10^0, respectively.
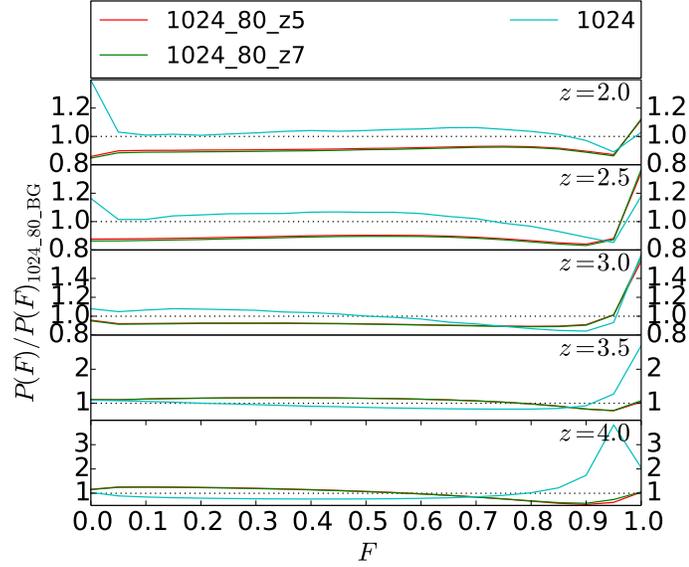
Figure 5.10: Ratios of H I Lyα flux PDF $P(F)$ between 3 simulations and 1024_80_BG at 5 redshifts. Simulation 1024 is a 1024^3, 20 Mpc/h, UV background-only simulation that is discussed in Chapter 3. Horizontal dotted lines are at ratio=1.

5.4 Flux power spectrum

With the same method described in Section 3.5, we calculate the flux power spectrum (FPS) of H I Lyman alpha forest absorption lines. We show the 1D power spectrum $P_{1D}(k)$ at $z = 2$ in Fig. 5.11. Our results are compared with observations from McDonald et al. (2005), Day (2013) linearly extrapolated to $z = 2$ and observations from Jena et al. (2005). It shows that with less absorptions in 1024_80_z5, its 1D FPS is lower than 1024_80_BG, and the discrepancy at large scale (i.e. lower value of k) is still significant. Results at other redshifts are shown in Fig. 5.12, where our values are still below observations. We get larger power spectrum in 1024_80_z5 than 1024_80_BG when $z > 3$, which is due to the patchiness in the He II reionization with quasars as point sources.
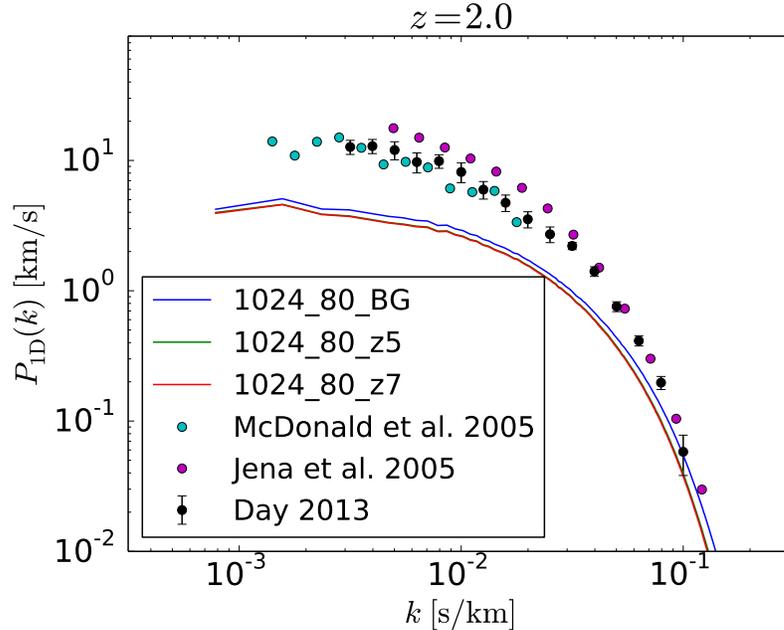
Figure 5.11: H I LAF 1D flux power spectrum $P_{1D}(k)$ at redshift $z = 2$. Observations from McDonald et al. (2005), Day (2013) linearly extrapolated to $z = 2$ and from Jena et al. (2005) are also shown.
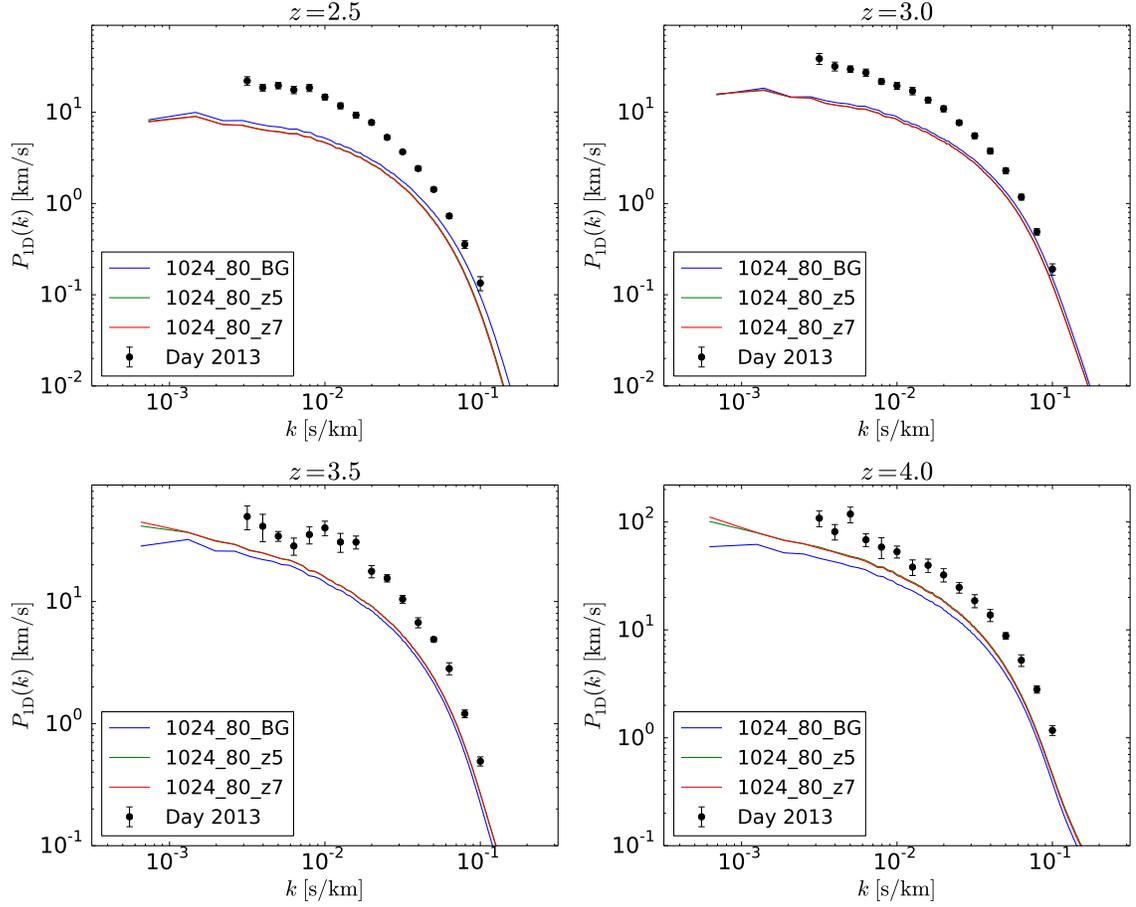
Figure 5.12: H I LAF 1D flux power spectrum $P_{1D}(k)$ at redshift $z = 2.5$ (Top Left), $z = 3$ (Top Right), $z = 3.5$ (Bottom Left), and $z = 4$ (Bottom Right). Observations from Day (2013) linearly extrapolated to the corresponding redshifts are also shown.
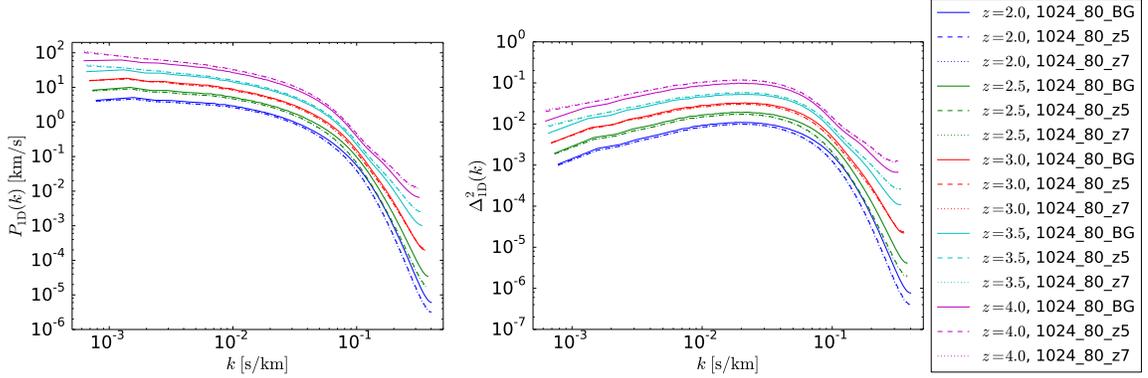
Figure 5.13: Left: 1D flux power spectrum $P_{1D}(k)$ of H I LAF at 5 redshifts. Right: The dimensionless power spectrum $\Delta^2_{1D}(k)$ at 5 redshifts. Left and right panel share the same labels. Lines with the same line style are from the same simulation, while lines with the same color are at the same redshift.
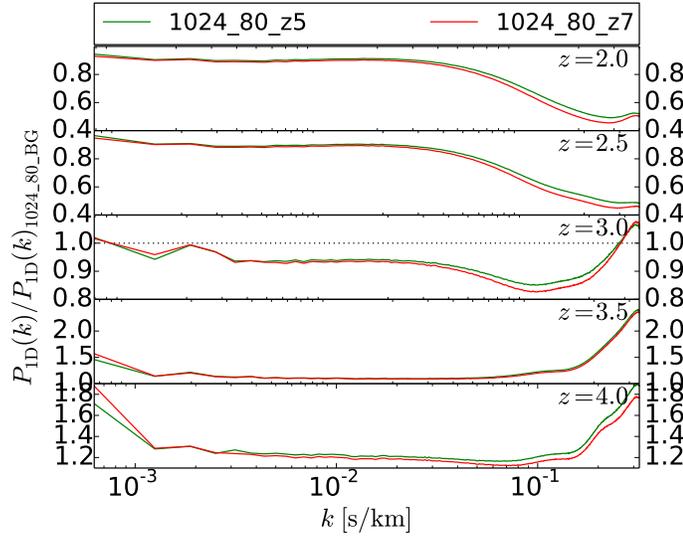
Figure 5.14: Ratios of H I Lyα flux power spectrum at 5 redshifts. The horizontal black dotted line at $z = 3.0$ is at ratio=1.

In Fig. 5.13 we show the 1D FPS $P_{1D}(k)$ (Left) and dimensionless FPS $\Delta_{1D}^2(k)$ (Right) at several redshifts. It shows that at higher redshifts the values of FPS at all scales are higher. We show ratios of FPS between 1024_80_z5 and 1024_80_BG at 5 redshifts in 5.14. According to the definition of dimensionless FPS, the ratio of two 1D FPS equals to the ratio of the two corresponding dimensionless FPS, so only one plot of ratios is shown. It shows that at $z > 3$ 1024_80_z5 has larger H I LAF power than 1024_80_BG in all scales, while it's reversed at $z < 3$.

5.5   2-point correlation functions

As another method to show the power of structures with repeated patterns, we calculate the 2-point correlation functions for temperature T (Fig. 5.15) and H I LAF flux F_{HI} (Fig. 5.16). The x-axis shows the velocity offset in km/s due to Hubble broadening of an increasing number of cells. A cell in both 1024_80_z5 and 1024_80_BG simulations is $\sim$78 kpc/h comoving. At $z = 2$ ($z = 4$) it corresponds to a Hubble broadening of $\Delta v = 7.8$ km/s (9.8 km/s). For each cell in one light ray or spectrum, we first calculate $\delta f(v) = f(v)/\bar{f} - 1$, then calculate the 2-point correlation function of f as $\xi(\Delta v) = \langle \delta f(v) \delta f(v + \Delta v) \rangle$, where f could be either T or F_{HI}, and the average is over all cells in $\sim 10^5$ random light rays and spectra.

The plots show that the fluctuations in T (indicated by points with $\Delta v = 0$) increase with time, while the fluctuations in F_{HI} decrease with time. At $z > 3$, fluctuations in T is larger in 1024_80_z5 than in 1024_80_BG. The relation reverses at lower redshifts. At $z < 3$, the 2-point correlation function of F_{HI} in 1024_80_z5 becomes larger than 1024_80_BG at $\Delta v >\sim 300$ km/s, which corresponds to ~ 3 Mpc/h comoving.
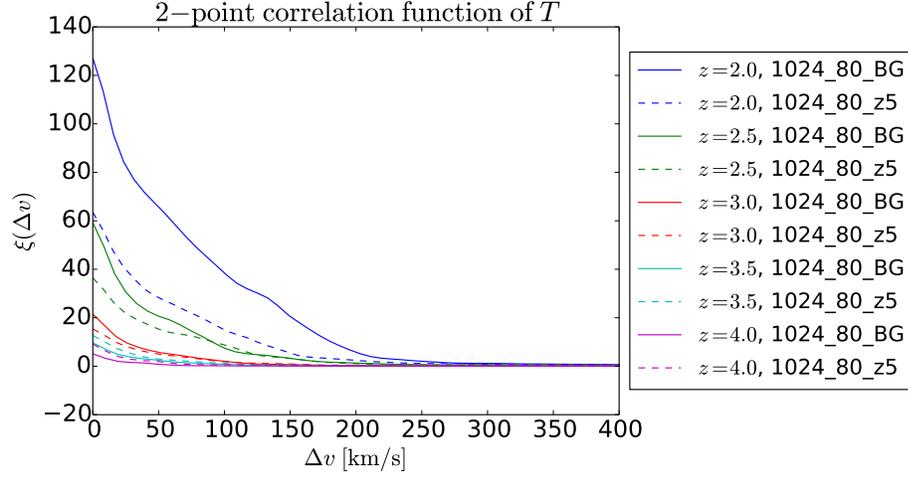
Figure 5.15: 2-point correlation functions of temperature T at 5 redshifts. The x-axis shows the Hubble broadening of an increasing number of cells.
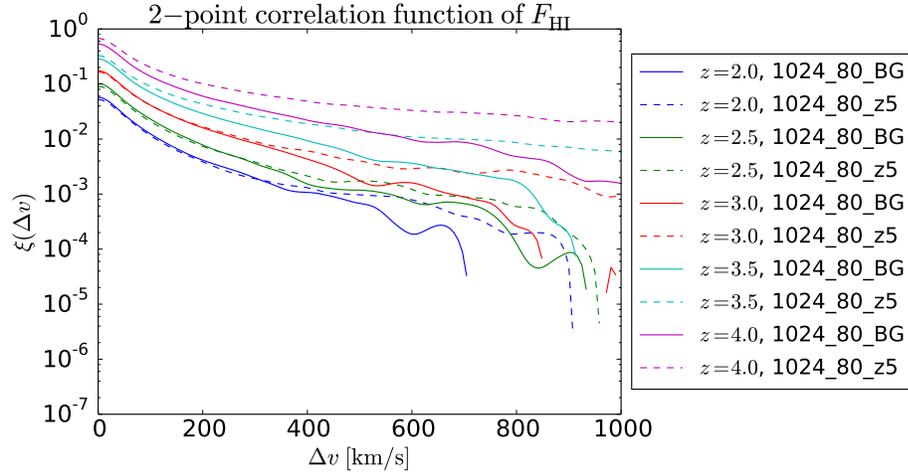
Figure 5.16: 2-point correlation functions of H I LAF flux F_{HI} at 5 redshifts. The x-axis shows the Hubble broadening of an increasing number of cells.

5.6   Doppler parameter

For each absorption spectrum, we fit the absorption lines with Voigt profiles using the tool AUTOVP (Davé et al., 1997) from which we obtain the column density

N and Doppler parameter b for each Voigt profile. We show their 2-dimensional histogram for 1024_80_z7 at $z = 2$ in Fig. 5.17.
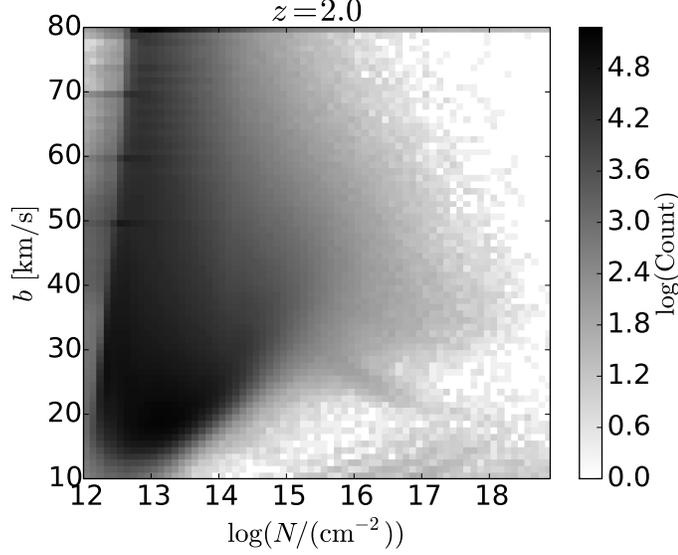
Figure 5.17: 2-d histogram of Doppler parameter b versus column density N for all 1024^2 spectra in simulation 1024_80_z7 at $z = 2$. We choose AUTOVP instead of yt to give better fitting results, but there are still visible artifacts, i.e. peaks in fitted b values.

Then for all absorbers with column density N satisfying $10^{12.5}/\text{cm}^2 < N < 10^{14.5}/\text{cm}^2$, we get the distribution of Doppler parameter b, and fit it with the Hui-Rutledge function (Eqn. 3.2) to get b_σ. We perform such analysis for 5 redshifts. The cell size of simulations 1024_80_BG, 1024_80_z5, and 1024_80_z7 is $L_{\text{cell}} \sim 80\text{kpc}/h$. As a result, the Hubble broadening of one cell is ~ 8 km/s at $z \sim 2$ and ~ 10 km/s at $z \sim 4$ (Fig. 3.9), making it hard to have accurate measurements of line widths, especially at high redshift. In fact, as redshift goes higher, the shape of the distribution of fitted Doppler parameters deviates from Hui-Rutledge function more and more. To reduce the defect of the limited resolution, at each redshift z, we calculate the ratio $b_{\sigma,2048}(z)/b_{\sigma,1024\_80\_BG}(z)$, and multiply it to $b_{\sigma,1024\_80\_BG}(z)$, $b_{\sigma,1024\_80\_z5}(z)$, and
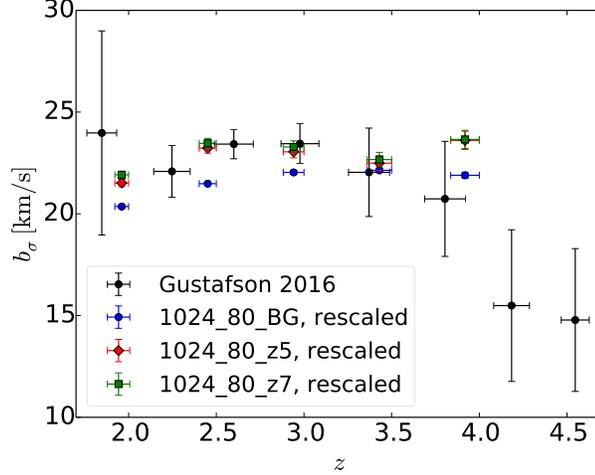
Figure 5.18: Fitted b_σ versus redshift. All simulated values are rescaled by multiplying the ratio between b_σ's from simulation 2048 and 1024_80_BG, in order to reduce the defects due to the large cell size. Data from Gustafson (2016) is shown for comparison. Our x-error bars occupy the range of redshift of light rays in that data dump. Our y-error bars are given by the fitting algorithm in the estimation of b_σ, then propagated according to rescaling formula.

$b_{\sigma,1024\_80\_z7}(z)$, respectively. Here simulation 2048 is discussed in Chapter 3. It is a 2048^3, 40 Mpc/h, UV background-only simulation, and has enough resolution to reach converged estimates of H I Lyα flux statistics (Lukić et al., 2015). The rescaled b_σ's are shown in Fig. 5.18. Data from Gustafson (2016) is shown for comparison. At $z < 3.5$ b_σ's in inhomogeneous He II reionization simulations (1024_80_z5 and 1024_80_z7) are higher than homogeneous optically thin simulation (1024_80_BG) by ~ 1 km/s, which is consistent with findings in Paschos et al. (2007). Since b_σ could be calculated from both observed and synthetic spectra and it's directly related with IGM temperature, it's a reliable indicator to compare thermal histories between observations and simulations. Overall our rescaled results agree with observations. Data from Gustafson (2016) shows a peak at $2.5 < z < 3$, so do our simulations. This peak happens close to the end of He II reionization (Fig. 4.5), and agrees with the

time when observed IGM temperature peaks (Fig. 4.7).

5.7 Distribution of column density of H I Lyα absorbers

Another property of each H I Lyα absorber is its column density N. To examine the distribution of H I Lyα absorbers versus column density N, we define the frequency distribution of H I absorbers per unit absorption distance as (Gustafson, 2016)

$$f(N) = \frac{\partial N_{\mathrm{ab}}}{X \partial N},$$ (5.1)

where $\mathrm{d}N_{\mathrm{ab}}$ is the number of H I Lyα absorbers with H I column density in the range $[N,\ N+\mathrm{d}N]$, and

$$X = \frac{H_0}{H(z)}(1+z)^2$$ (5.2)

is the absorption distance. We show $f(N)$ of 1024_80_z7 at $z = 2$ in Fig. 5.19. We then fit two power laws to the data. The first one is for absorbers with $13.3 < \log(N/(\mathrm{cm}^{-2})) < 16$, while the second one is for absorbers with $17 < \log(N/(\mathrm{cm}^{-2})) < 19.5$. We get two power-law slopes β_1 and β_2 for those two ranges. Due to the way $f(N)$ is defined, the comparison of the absolute value of $f(N)$ is meaningless, so we compare slopes β_1 and β_2 instead. We do such analysis for 5 redshifts, and the slopes are shown in Fig. 5.20. Data from Gustafson (2016) is shown for comparison.
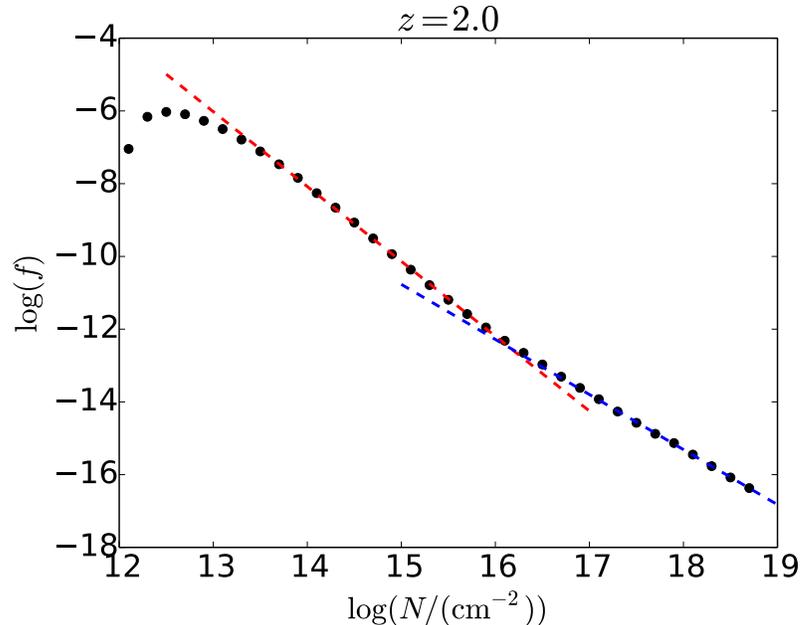
Figure 5.19: The frequency distribution of H I Lyα absorbers f in 1024_80_z7 at $z = 2$. We fit two power laws to the data. The first one for absorbers with $13.3 < \log(N/(\mathrm{cm}^{-2})) < 16$ has a slope $\beta_1 = -2.01 \pm 0.01$, while the second one for absorbers with $17 < \log(N/(\mathrm{cm}^{-2})) < 19.5$ has a slope $\beta_2 = -1.54 \pm 0.01$. Fitting results are shown as two dashed lines.
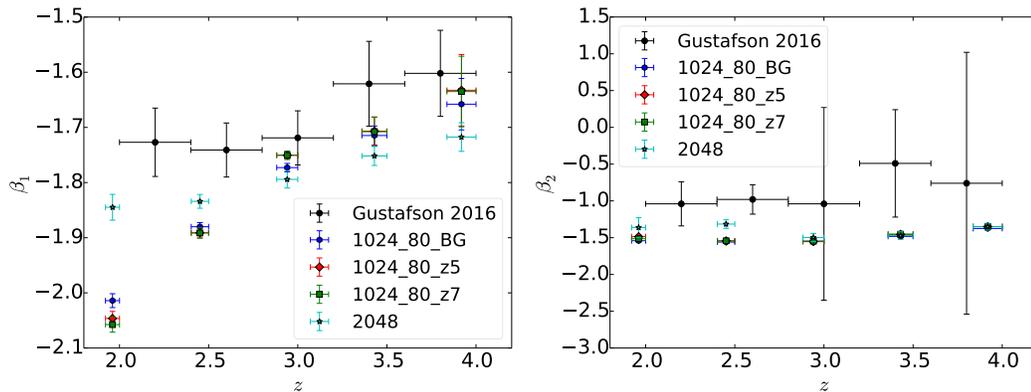
Figure 5.20: The power-law slopes in the frequency distribution of H I Lyα absorbers β_1 and β_2 versus redshift. Our x-error bars occupy the range of redshift of light rays in that data dump. Our y-error bars are given by the fitting algorithm in the estimation of β_1 and β_2. Data from Gustafson (2016) is shown for comparison. Note that their β_2 is fitted for absorbers with $17 < \log(N/(\mathrm{cm}^{-2})) < 21$, comparing with our range $17 < \log(N/(\mathrm{cm}^{-2})) < 19.5$.

5.8   Discussion

5.8.1   Effect of high z quasars

Here we compare our H I LAF spectral analysis results in 1024_80_z5 to 1024_80_z7. We don't do spectral analysis to 1024_126_z7 due to its low resolution.

Fig. 5.1 shows similar flux, H I number density, and temperature along the same line of sight for 1024_80_z5 and 1024_80_z7 from $z = 2$ to $z = 1.92$. In the similar plot shown a longer light ray from $z = 4$ to $z = 2$ (Fig. 5.3), there are several locations shown different temperatures (for example, $z \sim 3.6$), which is due to the random luminosity of quasars. Fig. 5.4 shows similar distribution of H I opacity for both simulations across 5 redshifts. The effective H I opacity at $z = 4$ is 0.996 for 1024_80_z5 and 0.983 for 1024_80_z7. The difference gets even smaller at lower redshifts (Fig. 5.5). They also have similar flux PDFs across all 5 redshifts (Fig. 5.7, 5.8, 5.9). 1024_80_z5 has slightly higher H I LAF flux power spectrum at small scales than 1024_80_z7 (Fig. 5.14). We conclude that high redshift quasars ($5 < z < 6.5$) do not affect the H I LAF spectral properties at $z \leq 4$ very much.

5.8.2   Discrepancies between observations and simulations of H I LAF

As mentioned in Section 1.4, there are currently two major discrepancies between observations and simulations of H I LAF, and it is the primary purpose of simulations discussed in this chapter to solve those discrepancies.

(1) The amplitude of the simulated 1D power spectrum is lower than observations. Specifically, the simulated power is 50% less than the observed one on large scales

$(\log(k/(\text{s/km})) < -2)$, and 20% less on small scales $-1.6 < \log(k/(\text{s/km})) < -1.1$ (Tytler et al., 2009) (Fig. 1.4 Left). With quasars as point sources, we are unable to reduce the discrepancy. In fact, our 1D power spectrum from 1024_80_z5 and 1024_80_z7 is slightly lower at all scales than 1024_80_BG and deviates further from the observed results (Fig. 5.11). There are several possible explanations we need to explore. (a) We have quasars as point sources, a thermal history better agreed with data, and a large box, but we do not have good resolution. A similar simulation with a better resolution (like 20 kpc/h comoving per cell) needs to the done. (b) As galaxies are the primary sources of H I reionization, we might also add galaxies as point sources to former simulations (like 1024_80_z7) from high redshifts until $z \sim 2$, using methods both in Chapter 4 and Chapter 7, and also use MGFLD for the radiation transfer. In this way we would capture the optically thick effects of both H I and He II reionization, thus an overall better thermal history. Even after the H I reionization is completed, the gas close to galaxies will still get extra ionization and heating. By replacing the H I ionizing background with points galaxies with multi-frequency radiation, we expect larger fluctuations in H I density fields. (c) Missing astrophysics, like pre-heating from X-rays, cosmic rays, and decaying particles might account for the discrepancy in power spectrum. (d) Something might be wrong with the current cosmological model.

(2) The H I LAF flux PDF has a wrong shape. Specifically, standard simulations got fewer pixels with a lot of absorption (e.g. flux = 0.15) and more pixels with less absorption (e.g. flux = 0.8) than data from Kim et al. (2007) (Fig. 1.4 Right). Flux PDFs from 1024_80_z5 and 1024_80_z7 agree well with data at $z = 2.0$ and $z = 2.5$, and show better agreement with data at $z = 3.0$ than 1024_80_BG (Fig. 5.7, 5.8). Even though the agreement is good, there are still unsolved problems:

(a) All simulated spectra has smaller number of pixels with flux close to zero than observations at all 3 redshifts. (b) The flux PDF from simulation 1024 (Fig. 3.10) shows similar discrepancy as other standard simulations (Fig. 1.4 Right), while our 1024_80_BG simulation shows much better agreement. So the better agreement might result from the larger simulation box or larger cell size instead of adding quasars as point sources. We still need simulations with a large box (like 80 Mpc/h comoving) and high resolution (like 20 kpc/h comoving per cell) to make it clear.

In next chapter we will generate artificial spectra of He II LAF and examine how quasars affect it as point sources.

Chapter 6

Lyman Alpha Forest simulations with Quasars. III. Spectroscopic Analysis of He II Lyα Forest

In this chapter we use similar methods to generate and analyze He II Lyα forest.

6.1   Synthetic He II absorption spectra

Fig. 6.1 shows one light ray from $z = 2.00$ to $z = 1.92$ (i.e. 80 Mpc/h comoving) and its normalized He II Lyα flux F, He II number density n_{HeII}, peculiar velocity along line of sight v_{los}, and temperature T. Each light ray has 1024 pixels, which is equal to our number of grids in one dimension. To see the effect of distributed sources, the same line of sight is shown for all simulations for comparison. It shows that typically 1024_80_z5 has higher temperature, less He II, and less absorption (i.e.
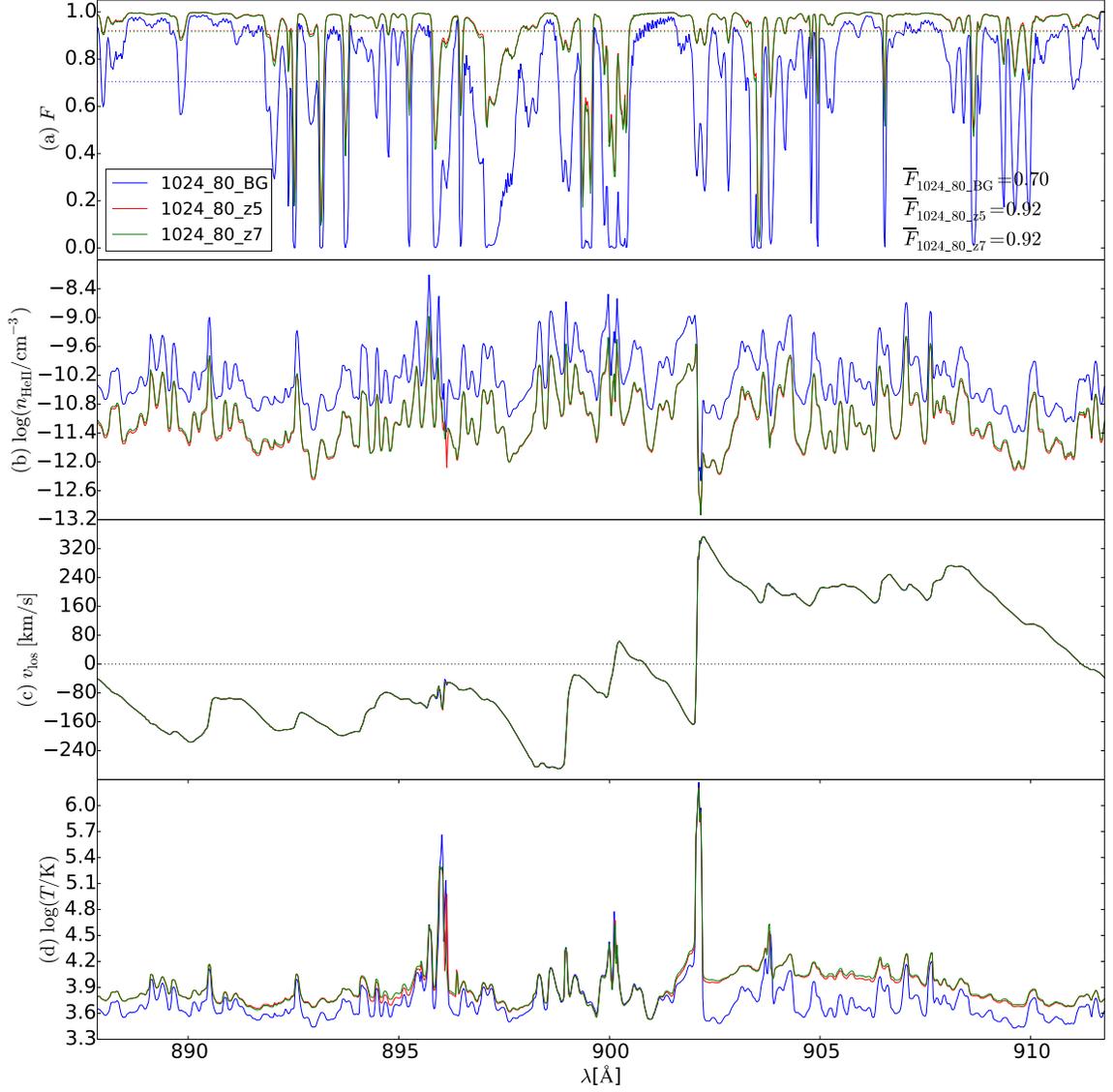
Figure 6.1: Quantities along a line of sight from $z = 2.00$ to $z = 1.92$. The same line of sight is used for all simulations shown. (a) Normalized flux of He II LAF. The average flux of this spectrum is shown at the lower right corner. (b) He II number density. (c) Peculiar velocity along line of sight. The black dotted horizontal line locates at $v_{\mathrm{los}} = 0$. (d) Temperature. Same color is used for one simulation across panels. In this line of sight 1024_80_z5 has higher temperature, less He II, and less absorption than 1024_80_BG.

larger normalized flux) comparing with 1024_80_BG. The peculiar velocities for both simulations are almost the same at each pixel. Fig. 6.2 shows another light ray from $z = 3.50$ to $z = 3.36$, where 1024_80_z5 shows more variations in temperature and He II number densities due to point quasars. This is to be expected as at this redshift the He II reionization is still patchy. A similar plot from $z = 4$ to $z = 2$ is generated with 22 data dumps (Fig. 6.3). It shows that at $z > 2.5$, in this line of sight 1024_80_z5 has lower temperature, more He II, and more absorption than 1024_80_BG, but they become closer later on.

For this line of sight there is virtually no transmitted flux for the portion of the spectrum $z > \sim 2.6$. The reason for this is the high density of He II along this section of the spectrum, due to incomplete reionization. A dip in the He II density and the corresponding patch of transmitted flux can be seen at $z \sim 3.13$. This is where the LOS intersects an ionized bubble of He III. The IGM is opaque to He II Ly α above this redshift. This contrasts with the spectrum shown in Fig. 6.2 derived from a different LOS which does exhibit some transmitted flux at higher redshifts. This points to the significant LOS to LOS variation that exists due to inhomogeneous He II reionization. We return to this topic later in this chapter.
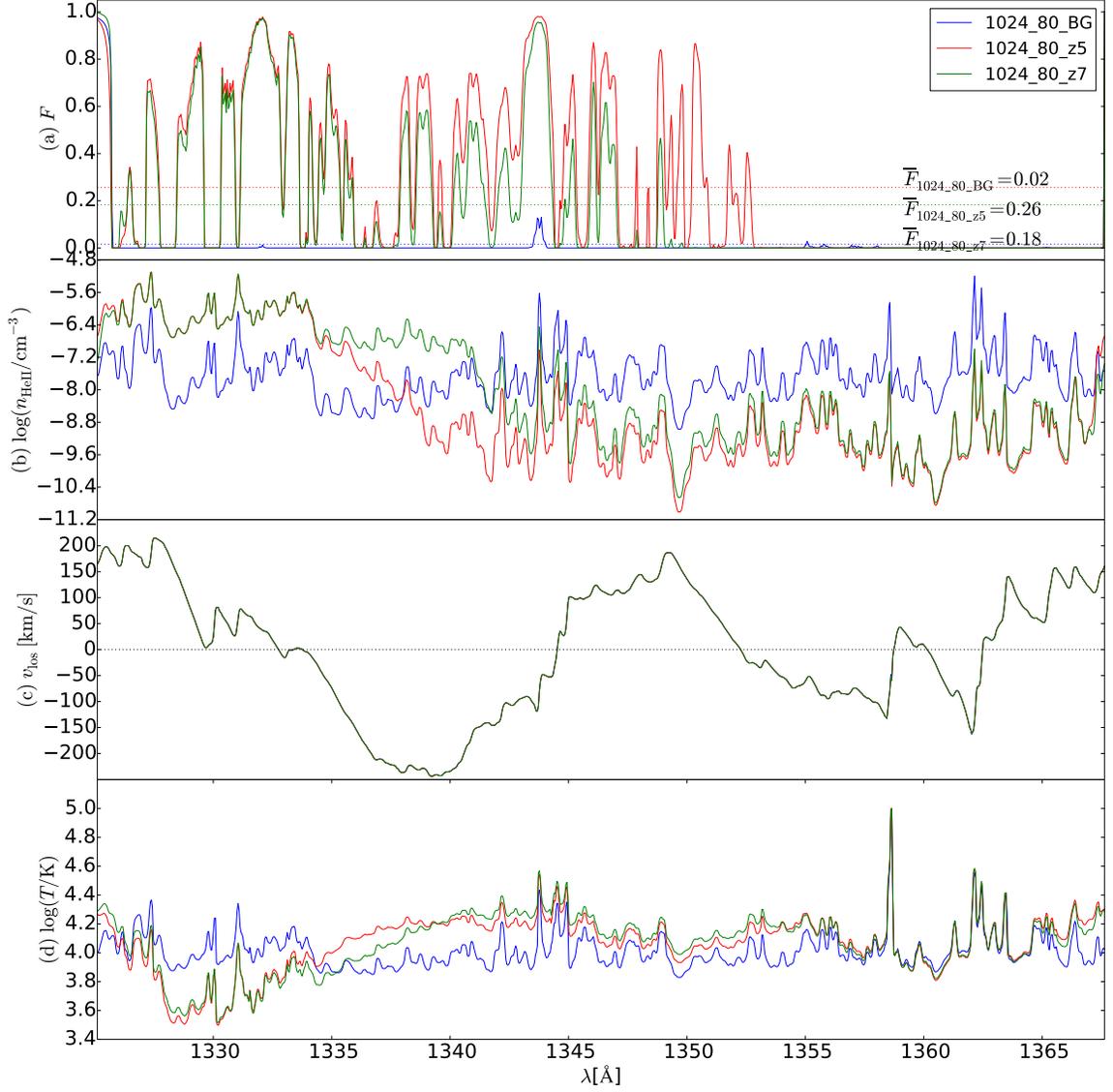
Figure 6.2: Quantities along a line of sight from $z = 3.5$ to $z = 3.36$. The same line of sight is used for all simulations shown. (a) Normalized flux of He II LAF. The average flux of this spectrum is shown at the lower right corner. (b) He II number density. (c) Peculiar velocity along line of sight. The black dotted horizontal line locates at $v_{los} = 0$. (d) Temperature. Same color is used for one simulation across panels. In this line of sight 1024_80_z5 shows more variations in temperature and He II number densities.
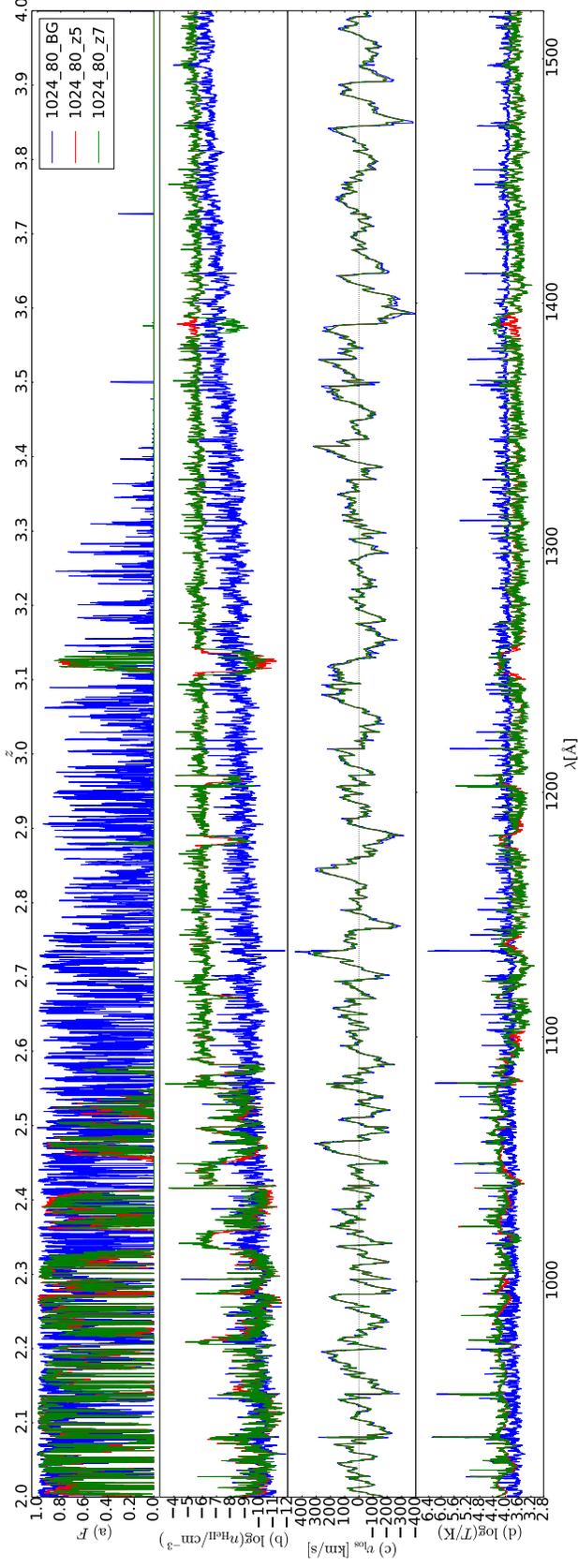
Figure 6.3: Quantities along a line of sight from $z = 4$ to $z = 2$. The same line of sight is used for all simulations shown. (a) Normalized flux. (b) He II number density. (c) Peculiar velocity along line of sight. The black dotted horizontal line locates at $v_{los} = 0$. (d) Temperature. Same color is used for one simulation across panels.

6.2 Optical depth
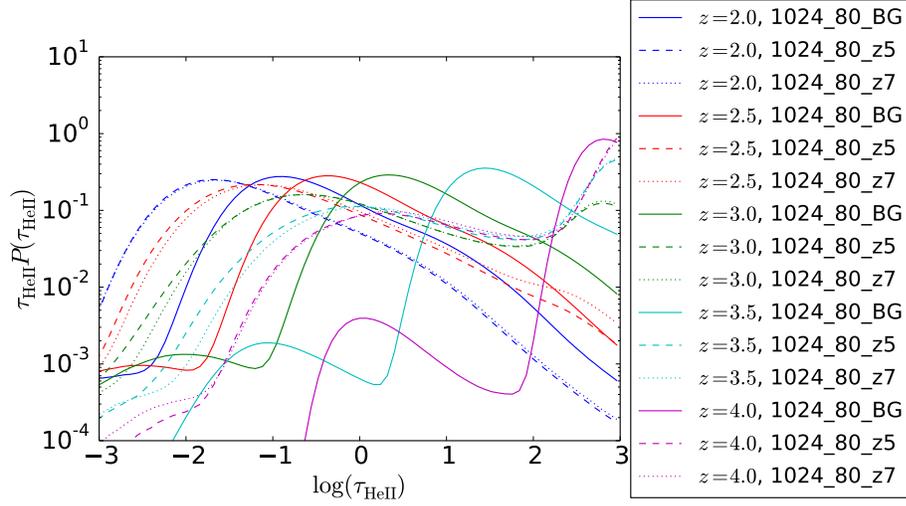
Figure 6.4: The distribution of He II opacity τ_{HeII} at different redshifts. Lines with the same line style are from the same simulation, while lines with the same color are at the same redshift.

The normalized flux F can be written as $F_i = e^{-\tau_i}$, where τ_i is the optical depth for pixel i. The distribution of He II opacity τ_{HeII} at different redshifts is shown in Fig. 6.4. It's different compared with the similar plot for H I (Fig. 5.4) in that most curves at high redshifts have two peaks, indicating that the He II reionization is still underway. At $z = 4$, 3.5 and 3, 1024_80_z5 has more cells with highest τ_{HeII}, more cells with lowest τ_{HeII}, and less cells in between, comparing with 1024_80_BG. This is due to the different ways we account for ionizing sources in those two simulations. With homogeneous ionization background in 1024_80_BG, every cell has the same He II ionization rate, even though the rate varies with time. But with point sources in 1024_80_z5, He II ionization rates are higher than average in cells close to sources, and lower than average in cells far from any sources at those high redshifts (see slices in the 4th row of Fig. 4.9), which results in the difference in the τ_{HeII} distribution.
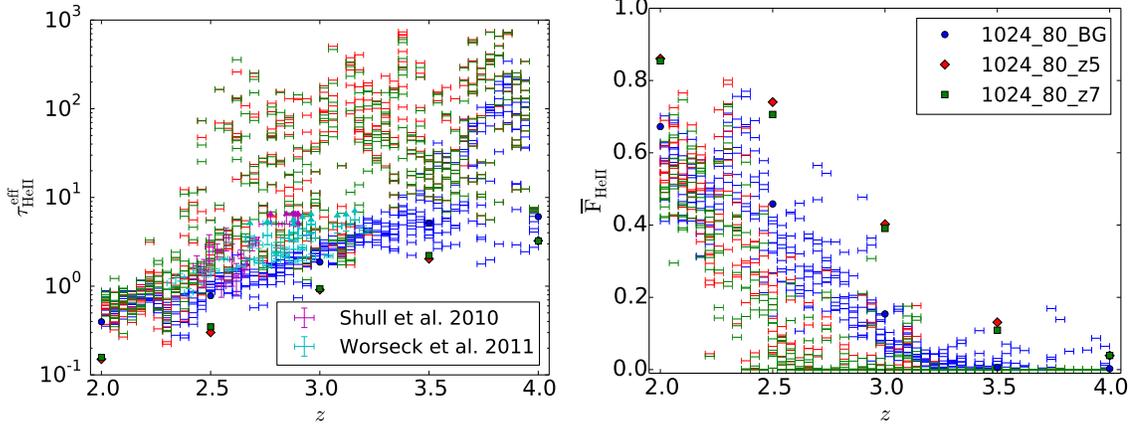
Figure 6.5: Effective He II opacity $\tau_{\mathrm{HeII}}^{\mathrm{eff}}$ (Left) and mean flux (Right) as a function of z. For each line of sight across the simulation box, we calculate its $\tau_{\mathrm{HeII}}^{\mathrm{eff}}$ from its mean flux. The average value over 1024^2 line of sights at certain redshifts are shown. We also show the sample values of 10 random light rays from $z = 4$ to $z = 2$. Labels for simulations are the same in the left and right panel. The color of sampled data is same as the simulation it comes from. In left panel observation data from Shull et al. (2010) and Worseck et al. (2011) is also shown for comparison.

At lower redshifts, $z = 2.5$ and 2, the percolation of He III bubbles has finished. The τ_{HeII} distributions have similar shape in 1024_80_z5 and 1024_80_BG, but He II is more ionized in 1024_80_z5 (see Fig. 4.11).

We also calculate the effective He II opacity $\tau_{\mathrm{HeII}}^{\mathrm{eff}}$, where τ^{eff} is defined as $e^{-\tau^{\mathrm{eff}}} = \langle F \rangle$. For each of 1024^2 line of sights across the simulation box at a certain redshift, we calculate its τ^{eff} from its mean flux. At 5 redshifts we do such analysis and the average values are shown in Fig. 6.5. Data from Shull et al. (2010) and Worseck et al. (2011) is also shown for comparison. At all 5 redshifts $\tau_{\mathrm{HeII}}^{\mathrm{eff}}$ in 1024_80_z5 is smaller than values in 1024_80_BG and data. To get an idea of fluctuations in $\tau_{\mathrm{HeII}}^{\mathrm{eff}}$, we generate ten long light rays from $z = 4$ to $z = 2$, and divide them into small bins with $\Delta z = 0.04$. Then we show their mean flux and effective optical depth in Fig. 6.5. At $z > 2.5$ there are large fluctuations in $\tau_{\mathrm{HeII}}^{\mathrm{eff}}$, which is consistent with observations
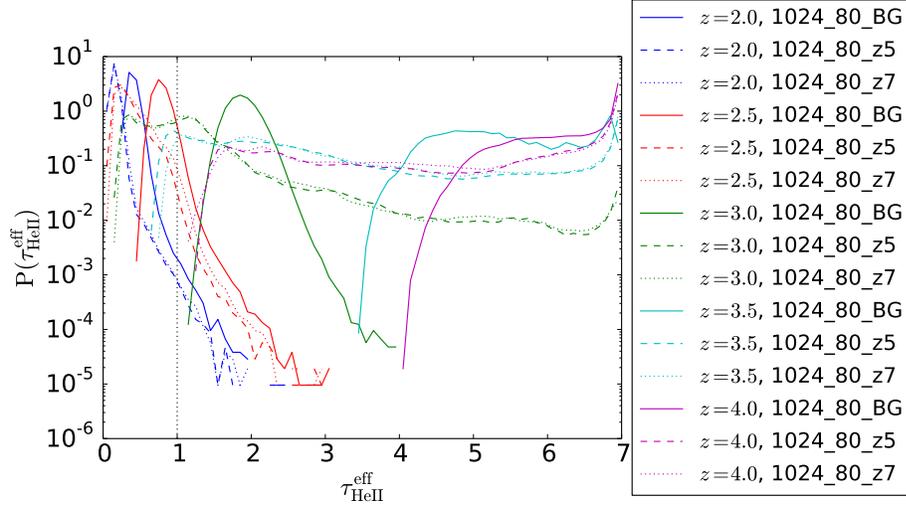
Figure 6.6: PDF of effective He II opacity $\tau_{\mathrm{HeII}}^{\mathrm{eff}}$ at redshift $z = 4, 3.5, 3, 2.5$, and 2. The vertical dotted line is at $\tau_{\mathrm{HeII}}^{\mathrm{eff}} = 1$. Lines with the same line style are from the same simulation, while lines with the same color are at the same redshift.

from Worseck et al. (2014) (Fig. 1.2).

Examining Fig. 6.5, we see that the LOS to LOS variation for $\tau_{\mathrm{HeII}}^{\mathrm{eff}}$ in the standard optically thin simulation 1024_80_BG is small for $z <\sim 3.2$, in conflict with the observations of Worseck et al. (2014). The inhomogeneous simulations 1024_80_z5 and 1024_80_z7 show large scatter, with $\tau_{\mathrm{HeII}}^{\mathrm{eff}}$ exceeding 100 for some LOS's. Given that most of Worseck et al.'s data points are lower limits, a direct comparison to data is difficult. The most we can say is our results are not inconsistent with the results of Worseck et al. (2014).

We show the probability distribution of $\tau_{\mathrm{HeII}}^{\mathrm{eff}}$ at 5 redshifts in Fig. 6.6. At $z \geq 3$ the shape of the distributions is quite different in 1024_80_BG and in 1024_80_z5. At all redshifts except $z = 3$, 1024_80_z5 has more lowest $\tau_{\mathrm{HeII}}^{\mathrm{eff}}$ and less highest $\tau_{\mathrm{HeII}}^{\mathrm{eff}}$ than 1024_80_BG. According to the convention to define $\tau^{eff} < 1$ as optically thin, all simulations are almost optically thin to He II ionizing photons below $z \sim 2.5$, which

means that an optically thin background treatment for He II ionization is not a good approximation to this problem above $z \sim 2.5$.

6.3   Flux PDF

With all the normalized flux F_i's, we have generated the flux probability distribution function (PDF) $P(F)$. Calculations of $P(F)$ follow the convention in Kim et al. (2007), where we divide the flux of all pixels into bins with width $\Delta F = 0.05$. Pixels with $F < 0.025$ or $F > 0.975$ are counted in the $F = 0$ bin and the $F = 1.0$ bin, respectively. The $P(F)$ of all 3 simulations at 5 redshifts are shown in Fig. 6.7. We show the ratios of $P(F)$ between 1024_80_z5 (1024_80_z7) and 1024_80_BG at 5 redshifts in Fig. 6.8.

6.4   Flux power spectrum

With the same method described in Section 3.5, we calculate the flux power spectrum (FPS) of the He II Lyman alpha forest absorption lines. In Fig. 6.9 we show the 1D FPS $P_{1D}(k)$ (Left) and dimensionless FPS $\Delta_{1D}^2(k)$ (Right) of the He II LAF at several redshifts. It shows that at all redshifts the values of the FPS at almost all scales are higher in 1024_80_BG than in 1024_80_z5.

We show ratios of FPS between 1024_80_z5 (1024_80_z7) and 1024_80_BG at 5 redshifts in Fig. 6.10. According to the definition of dimensionless FPS, the ratio of two 1D FPS equals to the ratio of the two corresponding dimensionless FPS, so only one plot of ratios is shown.
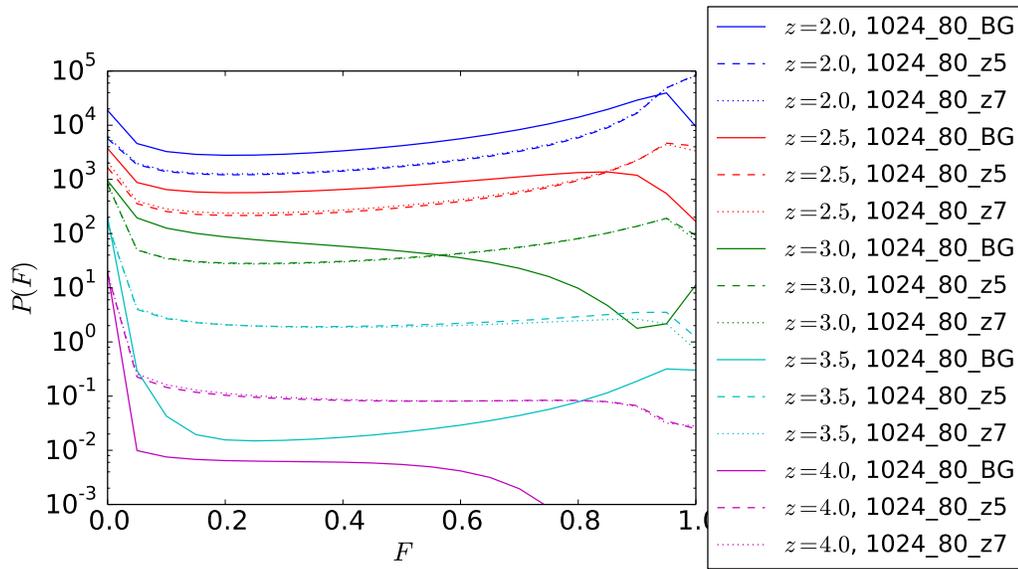
Figure 6.7: PDF of He II Lyα flux $P(F)$ at 5 redshifts. Lines with the same line style are from the same simulation, while lines with the same color are at the same redshift. From top to bottom, we show $P(F)$ at $z = 2$, 2.5, 3, 3.5, 4, and they are multiplied by a factor of 10^4, 10^3, 10^2, 10^1, 10^0, respectively.
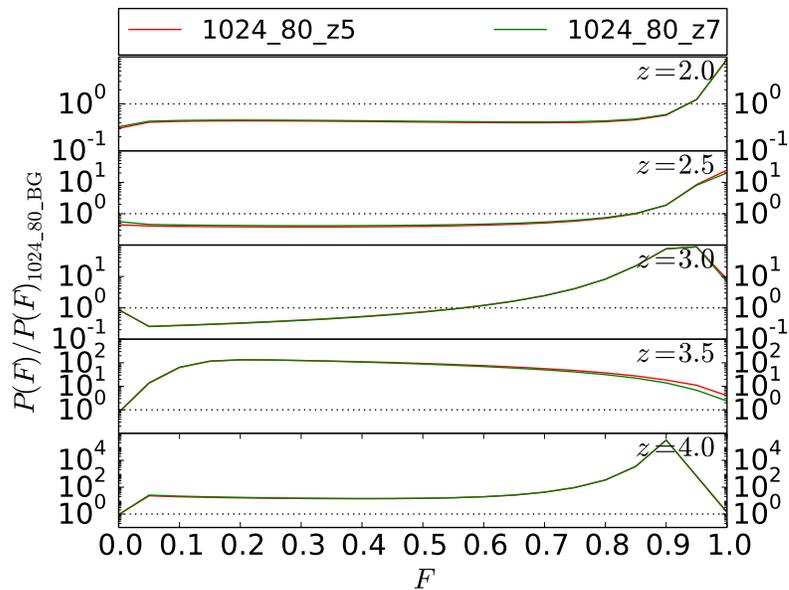
Figure 6.8: Ratios of He II Lyα flux PDF $P(F)$ between 1024_80_z5 (1024_80_z7) and 1024_80_BG at 5 redshifts. Horizontal dotted lines are at ratio=1.
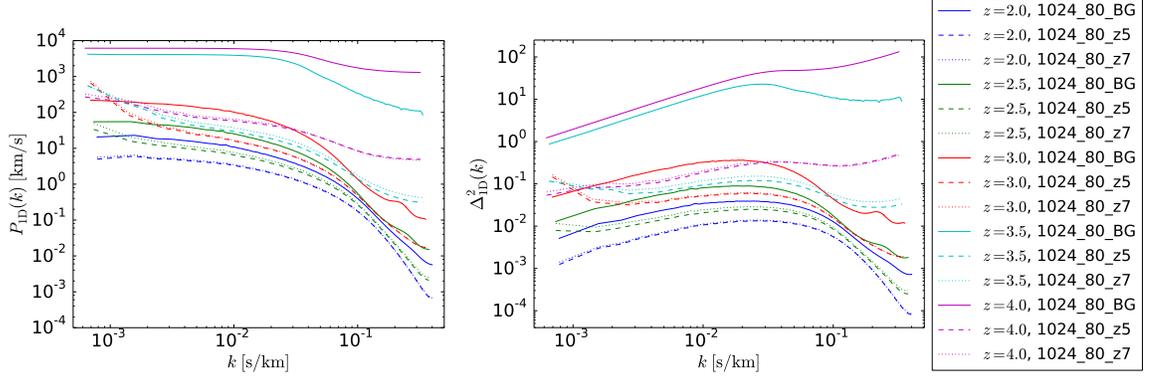
Figure 6.9: Left: 1D flux power spectrum $P_{1D}(k)$ of He II LAF at 5 redshifts. Right: The dimensionless power spectrum $\Delta^2_{1D}(k)$ at 5 redshifts. Left and right panel share the same labels. Lines with the same line style are from the same simulation, while lines with the same color are at the same redshift.
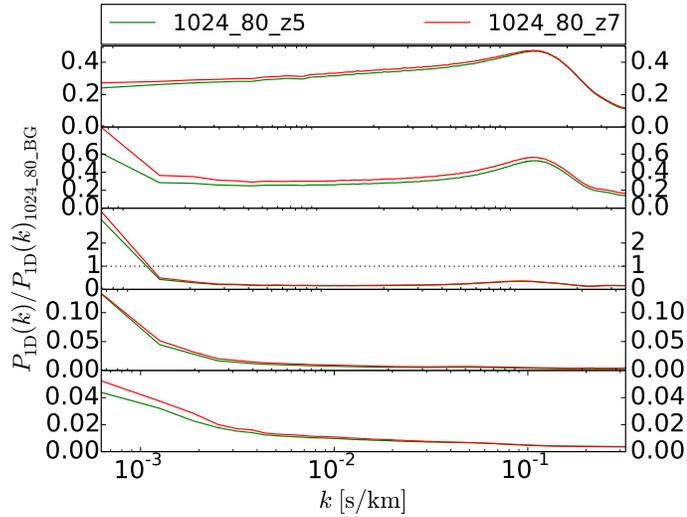
Figure 6.10: Ratios of He II Lyα flux power spectrum at 5 redshifts. The horizontal black dotted line at $z = 3.0$ is at ratio=1.

6.5 Effect of high z quasars

To examine the influence of high redshift QSOs on the spectroscopic properties of the He II LAF, here we compare simulations 1024_80_z5 and 1024_80_z7 which differ only the inclusion of faint, high z QSOs (Chap. 4). We do not perform a spectroscopic analysis of the larger box simulations 1024_126_z5 and 1024_126_z7 due to their lower spatial and hence spectroscopic resolution.

Fig. 6.1 shows similar flux, He II number density, and temperature along the same line of sight for 1024_80_z5 and 1024_80_z7 from $z = 2$ to $z = 1.92$. In the similar plot shown a longer light ray from $z = 4$ to $z = 2$ (Fig. 6.3), there are several locations shown different temperatures (for example, $z \sim 3.6$), which is due to the random luminosity of quasars. Fig. 6.4 shows similar distribution of He II opacity for both simulations across 5 redshifts. The effective He II opacity is also close (Fig. 5.5). They also have similar flux PDFs (Fig. 6.7) and flux power spectrum (Fig. 6.10) across all 5 redshifts. We conclude that high redshift quasars ($5 < z < 6.5$) do not affect the He II LAF spectral properties at $z \leq 4$ very much.

Chapter 7

Fully Coupled Simulation of Cosmic Reionization. III. Contribution of the Smallest Galaxies

7.1 Abstract

In Wise et al. (2014) we identified a new class of early galaxy that we estimate contributes up to 30% of the ionizing photons responsible for reionization. These are low mass halos in the range $M_{vir} = 10^{6.5} - 10^8 M_\odot$ that have been chemically enriched by supernova ejecta from prior Pop III star formation. Metal line cooling allows them to form stars even though their virial temperature is below the threshold for efficient cooling by atomic hydrogen. Despite their low star formation rates, these Metal Cooling halos (MCs) are significant sources of ionizing radiation, especially at the

onset of reionization, due to their high number density and ionizing escape fractions. Here we present a fully-coupled radiation hydrodynamic simulation of reionization that includes these MCs as well the more massive hydrogen atomic line cooling halos. Our method is novel: we perform halo finding inline with the radiation hydrodynamical simulation, and assign escaping ionizing fluxes to halos using a probability distribution function measured from the ultra-high resolution *Renaissance Simulations* (Xu et al. (2016a)). The PDF captures the mass dependence of the ionizing escape fraction as well as the probability that a halo is actively forming stars at a given time. With the inclusion of MCs, reionization starts earlier than if only halos of $10^8 M_\odot$ and above are included, however the redshift when reionization completes is only marginally affected, as this is driven by more massive galaxies. Because star formation is intermittent in MCs, the earliest phases of reionization exhibits a stochastic nature, with small H II regions forming and recombining. Only later, after the characteristic halo mass scale has reached $\sim 10^9 M_\odot$, does reionization proceed smoothly in the usual manner deduced from previous studies. With no adjustable model parameters, our 1152^3 simulation in a 14.4 Mpc box begins reionizing at $z = 20$, is 10% ionized at $z = 10$, and fully ionizes at $z_{ov} = 7.1$. Although our box is not large enough to be statistically representative, our simulation is marginally consistent with the latest measurement of the electron scattering optical depth τ_{es} to the CMB. Our results suggest that the chemical contribution of Pop III stars may be more important than their ionizing contribution to τ_{es}.

7.2   Introduction

The relative contributions to reionization from halos in different mass ranges are still not clear. A useful taxonomy for discussion we follow here was introduced by Iliev et al. (2007). The halos hosting early galaxies could be divided into three categories according to their mass. The first are minihalos (MHs, $M_{\mathrm{halo}} < 10^8 M_\odot$), which host the formation of Population III stars but otherwise are not thought to be efficient star formers due to their low virial temperatures and low H_2 cooling efficiency. The second are low-mass atomic-cooling halos (LMACHs, $10^8 M_\odot < M_{\mathrm{halo}} < 10^9 M_\odot$), which have virial temperatures just above the threshold to excite H atomic line cooling and form stars inefficiently. The third are high-mass atomic cooling halos (HMACHs, $M_{\mathrm{halo}} > 10^9 M_\odot$), which cool and form stars more efficiently than the LMACHs.

Some work has been done to study the role of LMACHs and MHs in reionization (Iliev et al., 2007; Choudhury et al., 2008; Shapiro et al., 2012; Iliev et al., 2012; Ahn et al., 2012; Wyithe and Loeb, 2013) which show that with the smallest galaxies included reionization begins earlier and the intergalactic electron-scattering optical depth τ_{es} is boosted. However these authors find the late phase of reionization is still dominated by HMACHs and the overlap redshift z_{ov} is not significantly affected. In these studies the galaxy properties are not simulated directly, but rather assumed using simple parameterized models which directly relate a halo's mass to its ionizing emissivity. For the smallest galaxies this relation is exceedingly uncertain due to a variety of complex physical processes. For example, the formation of the smallest galaxies is possibly suppressed due to the large Lyman-Werner background which photodissociates the primary coolant H_2 (Ahn et al., 2012), and due to supernova feedback which depletes the halo of gas (Wyithe and Loeb, 2013). Some simulations are

used to predict the signatures of reionization on the high redshift 21cm background, and to discuss how 21cm observations could help to distinguish the relative contributions of galaxies of different masses to reionization (Shapiro et al., 2012; Iliev et al., 2012).

However, recently Wise et al. (2014) have shown using AMR radiation hydrodynamic simulations that minihalos that have been chemically enriched by supernova ejecta from prior Pop III or Pop II star formation can cool and form stars and significantly contribute to the overall ionizing photon budget of reionization. We refer to this new class of halos as metal-line cooling halos, or MCs. Follow-on simulations (the *Renaissance Simulations*) in much larger volumes by Xu et al. (2016a) provides the star formation rates (SFR), intermittency, and ionizing escape fractions in the MCs, LMACHs, and HMACHs with extremely high resolution and good statistics. Using these results as our input, we revisit the problem: what is the role of the lowest mass halos in the reionization process? The main improvement of our work compared with previous work is that the simulations shown here are fully coupled cosmological radiation hydrodynamic simulations, with a time-dependent treatment of the ionization kinetics, and emissivities assigned to the source halos dynamically, considering the intermittency of the contribution from MCs.

We find that because star formation is intermittent in MCs, the earliest phases of reionization exhibits a stochastic nature, with small H II regions forming and recombining. Only later, after the characteristic halo mass scale has reached $\sim 10^9 M_\odot$, does reionization proceed smoothly in the usual manner deduced from earlier studies. With no adjustable model parameters, our 1152^3 simulation in a 14.4 Mpc box begins reionizing at $z = 20$, is 10% ionized at $z = 10$, and fully ionizes at $z = 7.3$. Although our box is not large enough to be statistically representative, our simulation

is marginally consistent with the latest measurement of the electron scattering optical depth to the CMB suggesting that the chemical contribution of Pop III stars may be more important than their ionizing contribution.

This paper is organized as follows. We summarize the relevant results from Xu et al. (2016a) in Section 2. The description of computational method and inputs to the simulations is provided in Section 3. We show results in Section 4 and offer discussion and conclusions in Section 5.

7.3   Ionizing Photons from the Smallest Galaxies

Unlike the previous papers in this series (So et al., 2014; Norman et al., 2015), where ionizing emissivities were calculated from a simple star formation/feedback recipe incorporated in the simulation itself, here we import results from much higher resolution simulations which calculate the escaping ionizing photons of high redshift galaxies directly. How this is done is described in Secs. 7.4.2 and 7.4.3. This approach enormously relaxes the spatial resolution requirement on the global reionization simulation and takes advantage of more precise simulation results. Uniform grids may be employed for the reionization simulation, however they must have sufficient mass and spatial resolution to accurately capture the halo population of importance. In addition, we are able to use moment methods for the radiation transport, which do not get bogged down as reionization completes as some ray tracing methods do (Norman et al., 2015).

We import the results of Xu et al. (2016a) who performed three high-resolution AMR simulations with simulated regions of different over-densities to study the abundance and escape fraction f_{esc} of smallest galaxies during reionization. The

so-called *Renaissance Simulations* (see also O'Shea et al. (2015)) include both Pop II and Pop III star formation and their radiative, mechanical, and chemical feedback. They find that for $M_{\mathrm{halo}} \lesssim 10^7 M_\odot$, the mean f_{esc} is between about 0.4 to 0.6, then f_{esc} declines smoothly over the mass range $10^7 M_\odot < M_{\mathrm{halo}} < 10^8 M_\odot$ to less than 0.05 for LMACHs ($10^8 M_\odot < M_{\mathrm{halo}} < 10^9 M_\odot$), then rises to about 0.1 to 0.2 for HMACHs($M_{\mathrm{halo}} > 10^9 M_\odot$) (cf. Figs. 15-17 in Xu et al. (2016a)). They also calculate the fraction of halos with active star formation as a function of M_{halo}, and find that the fraction is less than 0.1 for $\sim 10^7 M_\odot$ halos, ~ 0.5 for $\sim 10^8 M_\odot$ halos and then 1.0 for halos larger than $10^{8.5} M_\odot$ (cf. Fig. 19 in Xu et al. (2016a)). With those results they derive the number of escaped ionizing photons per second $\dot{N}_{\mathrm{ion,esc}}$ as a function of M_{halo}. For smaller halos ($10^7 M_\odot < M_{\mathrm{halo}} < 10^9 M_\odot$); i.e., the MCs and LMACHs, there is a large scatter in the gas fraction, the star formation efficiency, the escape fraction f_{esc} and thus $\dot{N}_{\mathrm{ion,esc}}$. So rather than try to fit this data the $\dot{N}_{\mathrm{ion,esc}} - M_{\mathrm{halo}}$ relation is reproduced as a table (Table 7.1, not published in Xu et al. (2016a)). The table shows the probability distribution of $\dot{N}_{\mathrm{ion,esc}}$ in different M_{halo} bins.

For HMACHs, the correlation is tighter and they derive a fit of this relation,

$$\log_{10}(\dot{N}_{\mathrm{ion,esc}}) = 36.033 + 1.675 \times \log_{10}(M_{\mathrm{halo}}), \qquad (7.1)$$

which we adopt here. Both Table 7.1 and Equation 7.1 are derived from their simulation of a normal-density region at redshift 12.5. Since they find that the galaxy properties are principally dependent on their mass, and are almost independent on their forming environment and redshift, we take the $\dot{N}_{\mathrm{ion,esc}} - M_{\mathrm{halo}}$ relation (Table 7.1 and Equation 7.1) as the input of our simulations during the whole reionization era.

Table 7.1: M_{halo} dependent escaped ionizing photons rate $\dot{N}_{\text{ion,esc}}$

	$10^{7.0}$	$10^{7.1}$	$10^{7.2}$	$10^{7.3}$	$10^{7.4}$	$10^{7.5}$	$10^{7.6}$	$10^{7.7}$	$10^{7.8}$	$10^{7.9}$
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
...	...	...	...	...	...	...	...	...	...	...
0.76	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.78	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.20e46
0.80	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	3.08e48
0.82	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	9.60e47	3.37e48
0.84	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	3.67e48	5.19e48
0.86	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	7.20e48	2.34e49
0.88	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.90e48	1.56e49	5.04e49
0.90	0.00	0.00	0.00	1.44e48	6.40e48	1.12e46	4.42e46	8.67e48	1.66e49	8.86e49
0.92	0.00	0.00	0.00	1.68e49	1.44e49	1.81e49	1.39e49	5.68e49	2.35e49	1.39e50
0.94	0.00	0.00	5.50e48	5.41e49	5.18e49	4.11e49	4.64e49	9.74e49	2.87e49	1.81e50
0.96	0.00	5.35e49	3.81e49	7.95e49	1.97e50	7.44e49	1.12e50	2.38e50	1.02e50	1.88e50
0.98	1.82e50	3.35e50	1.69e50	2.43e51	5.21e50	3.49e50	3.00e50	3.71e50	1.67e50	7.50e50

	$10^{8.0}$	$10^{8.1}$	$10^{8.2}$	$10^{8.3}$	$10^{8.4}$	$10^{8.5}$	$10^{8.6}$	$10^{8.7}$	$10^{8.8}$	$10^{8.9}$
0.00	0.00	0.00	0.00	0.00	0.00	0.00	9.30e48	8.15e49	4.91e49	1.12e50
0.02	0.00	0.00	0.00	0.00	0.00	0.00	9.30e48	8.15e49	4.91e49	1.12e50
0.04	0.00	0.00	0.00	0.00	0.00	0.00	9.30e48	8.15e49	4.91e49	1.12e50
0.06	0.00	0.00	0.00	0.00	0.00	0.00	9.30e48	8.15e49	4.91e49	1.12e50
0.08	0.00	0.00	0.00	0.00	0.00	0.00	9.30e48	8.15e49	4.91e49	1.12e50
0.10	0.00	0.00	0.00	0.00	0.00	0.00	9.30e48	8.15e49	4.91e49	1.12e50
0.12	0.00	0.00	0.00	0.00	0.00	0.00	9.30e48	8.15e49	4.91e49	1.12e50
0.14	0.00	0.00	0.00	0.00	0.00	0.00	9.30e48	8.15e49	4.91e49	1.12e50
0.16	0.00	0.00	0.00	0.00	0.00	0.00	9.30e48	1.64e50	4.91e49	1.12e50

Table 7.1: M_{halo} dependent escaped ionizing photons rate $\dot{N}_{\mathrm{ion,esc}}$, **Continued**

0.18	0.00	0.00	0.00	0.00	0.00	0.00	9.30e48	1.64e50	4.91e49	1.12e50
0.20	0.00	0.00	0.00	0.00	1.45e47	0.00	9.30e48	1.64e50	4.91e49	1.12e50
0.22	0.00	0.00	0.00	0.00	1.45e47	0.00	9.30e48	1.64e50	4.91e49	1.12e50
0.24	0.00	0.00	0.00	0.00	1.45e47	0.00	9.83e48	1.64e50	4.91e49	1.12e50
0.26	0.00	0.00	0.00	0.00	1.45e47	0.00	9.83e48	1.64e50	4.91e49	1.12e50
0.28	0.00	0.00	0.00	0.00	1.45e47	0.00	9.83e48	1.64e50	4.91e49	1.12e50
0.30	0.00	0.00	0.00	0.00	1.45e47	0.00	9.83e48	1.64e50	4.91e49	1.12e50
0.32	0.00	0.00	0.00	0.00	1.45e47	0.00	9.83e48	3.92e50	6.88e50	1.12e50
0.34	0.00	0.00	0.00	0.00	1.45e47	0.00	9.83e48	3.92e50	6.88e50	1.12e50
0.36	0.00	0.00	0.00	0.00	1.45e47	0.00	9.83e48	3.92e50	6.88e50	1.12e50
0.38	0.00	0.00	0.00	0.00	1.45e47	0.00	9.83e48	3.92e50	6.88e50	1.12e50
0.40	0.00	0.00	0.00	0.00	6.38e48	0.00	9.83e48	3.92e50	6.88e50	1.12e50
0.42	0.00	0.00	0.00	0.00	6.38e48	0.00	9.83e48	3.92e50	6.88e50	1.12e50
0.44	0.00	0.00	0.00	0.00	6.38e48	0.00	9.83e48	3.92e50	6.88e50	1.12e50
0.46	0.00	0.00	0.00	0.00	6.38e48	0.00	9.83e48	3.92e50	6.88e50	1.12e50
0.48	0.00	0.00	0.00	0.00	6.38e48	0.00	9.83e48	3.92e50	6.88e50	1.12e50
0.50	0.00	0.00	0.00	0.00	6.38e48	0.00	1.16e49	5.39e50	6.88e50	1.54e51
0.52	0.00	0.00	0.00	0.00	6.38e48	0.00	1.16e49	5.39e50	6.88e50	1.54e51
0.54	0.00	0.00	0.00	0.00	6.38e48	0.00	1.16e49	5.39e50	6.88e50	1.54e51
0.56	0.00	0.00	0.00	0.00	6.38e48	1.43e48	1.16e49	5.39e50	6.88e50	1.54e51
0.58	0.00	0.00	0.00	1.39e47	6.38e48	1.43e48	1.16e49	5.39e50	6.88e50	1.54e51
0.60	0.00	0.00	2.60e47	1.39e47	9.03e49	1.43e48	1.16e49	5.39e50	6.88e50	1.54e51
0.62	0.00	0.00	2.60e47	1.39e47	9.03e49	1.43e48	1.16e49	5.39e50	6.88e50	1.54e51
0.64	0.00	2.22e47	9.04e47	1.60e48	9.03e49	1.43e48	1.16e49	5.39e50	6.88e50	1.54e51
0.66	0.00	2.22e47	9.04e47	1.60e48	9.03e49	1.43e48	1.16e49	1.12e51	6.96e50	1.54e51
0.68	0.00	2.22e47	9.04e47	1.60e48	9.03e49	1.43e48	1.16e49	1.12e51	6.96e50	1.54e51
0.70	0.00	3.08e48	1.57e48	8.42e48	9.03e49	2.08e50	1.16e49	1.12e51	6.96e50	1.54e51

Table 7.1: $M_{\rm halo}$ dependent escaped ionizing photons rate $\dot{N}_{\rm ion,esc}$, **Continued**

0.72	0.00	3.08e48	1.57e48	8.42e48	9.03e49	2.08e50	1.16e49	1.12e51	6.96e50	1.54e51
0.74	0.00	5.77e48	5.15e48	8.42e48	9.03e49	2.08e50	5.88e50	1.12e51	6.96e50	1.54e51
0.76	0.00	5.77e48	5.15e48	1.54e49	9.03e49	2.08e50	5.88e50	1.12e51	6.96e50	1.54e51
0.78	3.72e47	5.77e48	5.15e48	1.54e49	9.03e49	2.08e50	5.88e50	1.12e51	6.96e50	1.54e51
0.80	2.06e48	1.53e49	7.22e48	1.54e49	3.78e50	2.08e50	5.88e50	1.12e51	6.96e50	1.54e51
0.82	2.06e48	1.53e49	7.22e48	4.17e49	3.78e50	2.08e50	5.88e50	1.25e51	6.96e50	1.54e51
0.84	1.31e49	4.39e49	4.03e49	4.17e49	3.78e50	3.87e50	5.88e50	1.25e51	6.96e50	1.54e51
0.86	1.48e49	4.39e49	4.03e49	4.17e49	3.78e50	3.87e50	5.88e50	1.25e51	6.96e50	1.54e51
0.88	1.48e49	4.39e49	4.03e49	1.78e50	3.78e50	3.87e50	5.88e50	1.25e51	6.96e50	1.54e51
0.90	1.75e49	1.60e50	1.37e50	1.78e50	3.78e50	3.87e50	5.88e50	1.25e51	6.96e50	1.54e51
0.92	3.07e49	1.60e50	1.37e50	1.78e50	3.78e50	3.87e50	5.88e50	1.25e51	6.96e50	1.54e51
0.94	3.07e49	2.62e50	1.81e50	2.68e50	3.78e50	3.87e50	5.88e50	1.25e51	6.96e50	1.54e51
0.96	7.87e49	2.62e50	1.81e50	2.68e50	3.78e50	3.87e50	5.88e50	1.25e51	6.96e50	1.54e51
0.98	7.87e49	2.62e50	1.81e50	2.68e50	3.78e50	3.87e50	5.88e50	1.25e51	6.96e50	1.54e51

Notes. $M_{\rm halo}$ is in the unit of [$M_\odot$], while $\dot{N}_{\rm ion,esc}$ is in the unit of [s^{-1}].

7.4   Numerical Methodology

7.4.1   Basic Model

All simulations presented in this paper are carried out using the publicly available Enzo code (Bryan et al., 2014b). We use its built-in implicit flux-limited diffusion (FLD) radiative transfer solver (Norman et al., 2015) for the transfer of ionizing photons, which are treated in the grey approximation. In Norman et al. (2015) we show that Enzo's FLD and MORAY ray tracing radiative transfer solver (Wise and Abel, 2011) give nearly identical ionization histories in a reionization test problem. We use FLD because it is much faster, especially as the volume becomes fully ionized. The ionizing sources are assumed to be low metallicity star forming galaxies in halos of mass $M_{halo} \geq 10^7 M_\odot$. Thus we ignore the radiative contribution of Pop III stars. We return to the impact of this assumption in the discussion section. The spectral energy distribution (SED) of the stellar radiation is the same as in So et al. (2014), which is the SED derived by Ricotti et al. (2002) for a $Z = 0.04 Z_\odot$ stellar population but truncated above 4 Ryd. Because we input into the simulation the number of escaping ionizing photons measured at the virial radius from the *Renaissance Simulations* (Sec. 2), we do not need to assume an ionizing escape fraction. We also only need to resolve the virial radii of the halos of importance, which greatly relaxes the spatial resolution requirement, but not the mass resolution (see below). A WMAP7 ΛCDM cosmological model is used: $\Omega_M = 0.27$, $\Omega_\Lambda = 0.73$, $\Omega_b = 0.047$, $h = 0.7$, $\sigma_8 = 0.82$, and $n = 0.95$, where the variables have the usual definitions. All simulations start from redshift 99 and run until the simulation volume is fully ionized.

In this paper we present three simulations differing only in mass and spatial

resolution and box size. All use inline halo finding and assign emissivities to halos as described in Secs. 7.4.2 and 7.4.3. Their properties are summarized in Table 7.2. The first two constitute a resolution study, which show the importance of including halos as small as $10^7 M_\odot$. 256_all is a 256^3 cell/particle simulation in a 6.4 Mpc box. This is the same as the test problem presented in So et al. (2014) and Norman et al. (2015), and has the same mass and spatial resolution as the science run analyzed in So et al. (2014). In that case the dark matter particle mass was chosen so that the halo mass function was complete above $M_{halo} = 10^8 M_\odot$. 512_all is a 512^3 cell/particle simulation in the same box. It has $2\times$ the spatial resolution and $8\times$ the mass resolution as the 256_all simulation. The halo mass function is complete to $\sim 10^7 M_\odot$, essential for including the MCs. These simulations are discussed in Sec. 7.5.1. The third, 1152_all simulation, is our science run. It is a 1152^3 cell/particle simulation in a 14.4 Mpc box. It has the same mass and spatial resolution as 512_all simulation except in a box 2.25 as large (11.4 times the volume). This simulation is discussed in Sec. 7.5.2.

7.4.2   Inline Halo Finding

The Enzo code has an embedded Python capability which allows the user to execute Python scripts which operate on Enzo's internal data structures as the code runs. As the `yt` toolkit (Turk et al., 2011b) is implemented in Python, many of `yt`'s analysis capabilities can be run inline with the computation. This includes the Parallel HOP halo finder (Skory et al., 2010), which we employ here. In all simulations in Table 7.2, we use embedded Python to call a script every 20 million years to do the following: (i) find halos and calculate their integral properties, including their virial masses; (ii) assign an emissivity to each halo according to its mass (Section 7.4.3), (iii) zero the

Table 7.2: List of simulations in Chapter 7

Simulation	L_{box} (Mpc)	$N_{cell} = N_p$	$m_p(M_\odot)$	$100m_p(M_\odot)$	$\Delta x(kpc)$	z_{ov}
256_all	6.4	256^3	4.8×10^5	4.8×10^7	25	5.9
512_all	6.4	512^3	6×10^4	6×10^6	12.5	6.0
1152_all	14.4	1152^3	6×10^4	6×10^6	12.5	7.1

Notes. Box size is in the unit of comoving Mpc.

old emissivity field array, and use the halo's positions and emissivities to compute a new emissivity field array (Section 7.4.3). We choose 20 million years since that is the typical lifetime of OB stars in a coeval stellar population. The emissivity fields are then kept constant until the next inline Python script is called 20 million years later. After clearing all the old emissivity fields, we distribute the new halo emissivity evenly into 27 adjacent cells (a 3×3×3 cube) centered at the cell that contains the halo's center of mass. Since $\dot{N}_{\rm ion,esc}$ (Equation 7.1 and Table 7.2) already include escaping fractions, by distributing the emissivity field in a larger region instead of the center cell we avoid the ionizing photons being absorbed again.

Because low mass halos ($10^7 M_\odot < M_{\rm halo} < 10^{8.5} M_\odot$) do not have a unit probability of actively forming stars (Xu et al. (2016a), Table 7.1), halos emitting during this 20 million years will likely not emit for the next 20 million years. In this way we take the intermittency of the contribution from low mass halos into consideration.

7.4.3 Assigning the Emissivity Field

Instead of using a constant mass-to-light ratio (Iliev et al., 2007), we assign emissivities to halos according to the $\dot{N}_{\rm ion,esc} - M_{\rm halo}$ relations from Xu et al. (2016a). In all simulations, for the larger LMACH halos ($M_{\rm halo} > 10^9 M_\odot$), we use Equation 7.1

to get their $\dot{N}_{\mathrm{ion,esc}}$. For each halo with mass between 10^7 to 10^9 $M_\odot$, we generate a random number between 0 and 1 and use it to choose the corresponding $\dot{N}_{\mathrm{ion,esc}}$ from Table 7.1 in its mass bin. Then we calculate the emissivity of that halo assuming that each ionizing photons is 21.6eV. Note that most of halos with mass below $10^{8.4} M_\odot$ have zero emissivity. For example, for halos with mass around $10^7 M_\odot$, there are only less than 2% of them with nonzero emissivity. This is due to the inefficiency in star formation and the supernova feedback in low mass halos (Wyithe and Loeb, 2013). Each halo larger than $10^{8.6} M_\odot$ has nonzero emissivity, but the value may change every 20 million years when a new value is chosen in its mass bin or when it falls into another mass bin.

7.5 Results

7.5.1 Resolution study $-$ 256^3 and 512^3 simulations

To understand the role of the smallest galaxies in reionization we did two simulations, 256_all and 512_all, both with 6.4 comoving Mpc per side. As described above, the 512_all simulation has twice the spatial resolution and eight times the mass resolution as the 256_all simulation. These two simulations thus constitute a small resolution study. The parameters were chosen so that the halo mass functions are complete to $10^8 M_\odot$ and $10^7 M_\odot$, respectively.

Figure 7.1 shows the halo counts in 3 mass bins in both simulations: MCs - $10^7 \leq M_{halo}/M_\odot \leq 10^8$; LMACHs - $10^8 \leq M_{halo}/M_\odot \leq 10^9$; and HMACHs - $M_{halo}/M_\odot > 10^9$. Referring to the 512_all curves, we see MCs, LMACHs, and HMACHs begin forming at $z \sim 18, 15, 11$, respectively. Comparing these curves to
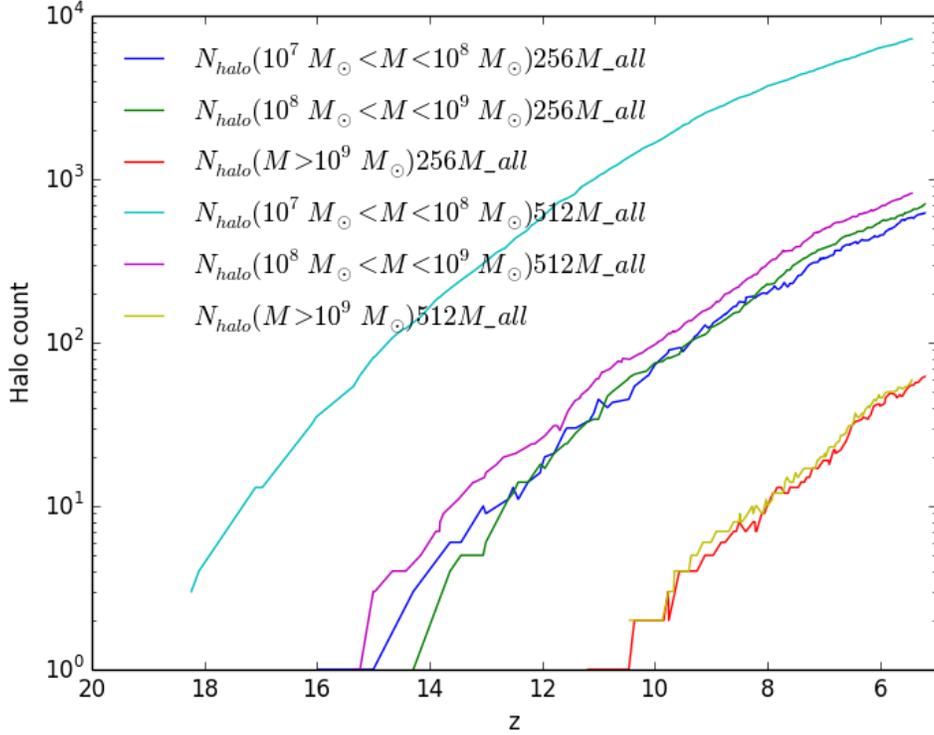
Figure 7.1: Halo counts vs. redshift for the 256_all and 512_all resolution study simulations for three mass bins: $10^7 \leq M_{halo}/M_{\odot} \leq 10^8$; $10^8 \leq M_{halo}/M_{\odot} \leq 10^9$; and $M_{halo}/M_{\odot} > 10^9$. Note the virtual absence of halos in the lowest mass bin corresponding to metal-line cooling halos (MCs) in the lower resolution simulation 256_all.

their 256_all counterparts, we see that the LMACH and HMACH formation histories are converged, but the MCs are severely underestimated. Due to the higher mass resolution, the MC halo counts in 512_all are 1.5 to 2 orders more than those in 256_all, and they begin forming sooner. The first halo with $M_{\rm halo} > 10^7 M_{\odot}$ appears at redshift $\sim$18.2 in 512_all, which is earlier than redshift $\sim$16.0 in 256_all.

Figure 7.2 shows the evolution of the ionizing photons emitted as a function of redshift in the 256_all and 512_all simulations. In the top panel we show the comoving ionizing luminosity density from halos below and above $M_{halo} = 10^9 M_{\odot}$, as well as the total; in the bottom panel we show the fraction of the total ionizing luminosity coming
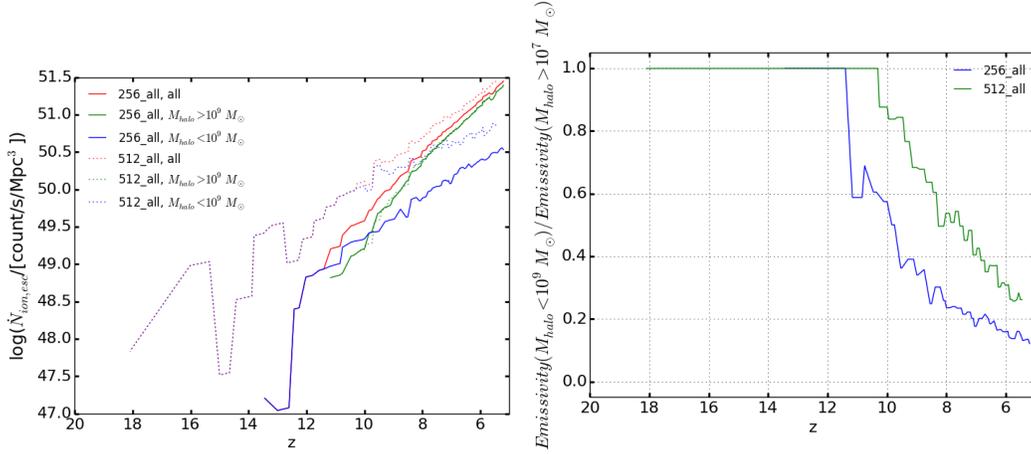
Figure 7.2: Contribution of halos of different masses to the ionizing emissivity as a function of redshift. Top: comoving ionizing luminosity density vs. redshift from halos below and above $10^9 M_\odot$ and their sum, for the 256_all test simulation (solid lines) and 512_all test simulation (dotted lines). Bottom: Fraction of ionizing photons from halos below $10^9 M_\odot$ versus redshift for 256_all and 512_all. The fraction drops as HMACHs become the dominant ionizing sources.

from halos below $M_{halo} = 10^9 M_\odot$. Looking at the top panel we can see that the HMACH contribution becomes dominant below $z \sim 10$ in the 256_all simulation, but not until $z \sim 8$ in the 512_all simulation . Because the HMACH population is virtually identical in both simulations, the difference is due to the enhanced contribution of the low mass halos, and specifically the MCs since the LMACH populations are also virtually identical in the two simulations (Fig. 7.1). Referring to the bottom panel, when there are no HMACHs, all the ionizing photons come from low mass halos so the ratio is one. Then as HMACHs form and become dominant the ratio drops to $\sim 15\%(25\%)$ when the reionization completes in 256_all(512_all), with some fluctuations in between. The ratio is always higher in 512_all because it has more low mass halos than 256_all but about the same amount of HMACHs.

Figure 7.3 shows the evolution of the volume fraction ionized above an ionization fraction of 10% for the 256_all and 512_all simulations. Although the first halo with
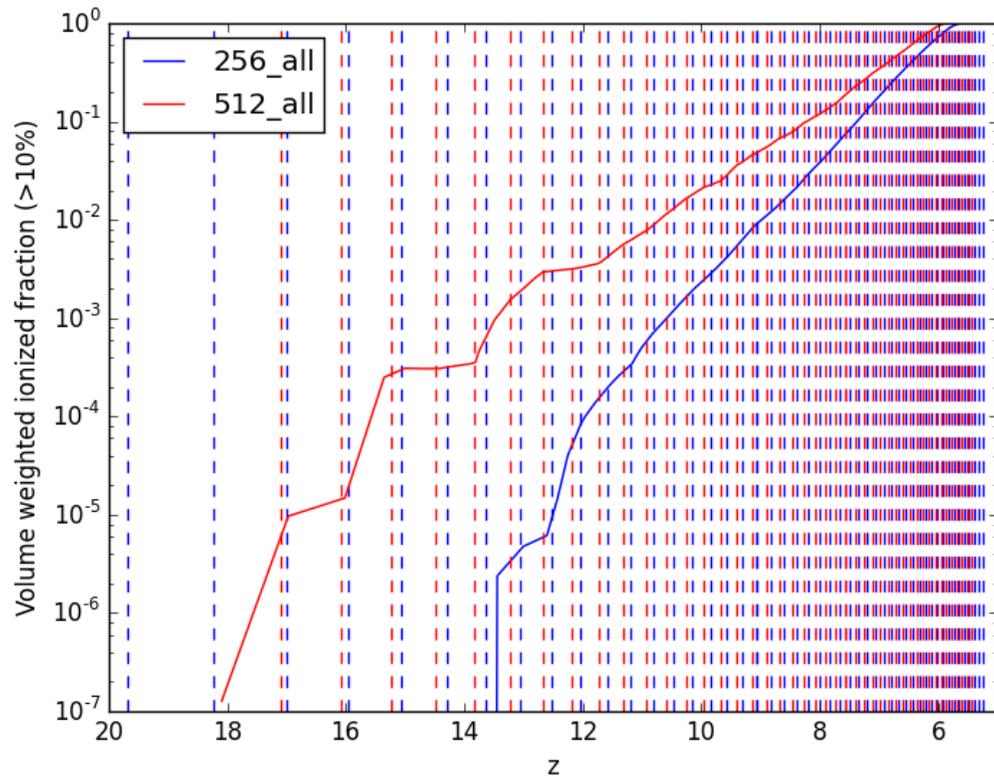
Figure 7.3: The ionized volume fraction as a function of redshift for 256_all and 512_all. The vertical dashed lines are approximately 20 million years apart indicating the time when inline Python works to find halos and assign ionizing emissivities according to the current halos' mass as described in Sec. 7.4.3.

$M_{\mathrm{halo}} > 10^7 M_\odot$ appears at redshift $\sim$16.0 in 256_all, the volume doesn't begin to ionize until redshift $\sim$13.5. This is due to the low probability for lower mass halos to emit (Table 7.1). In 512_all the time between first low mass halo emits and the first HMACH emits is longer than in 256_all, and there are several "stair steps". This is also due to the randomness in the turning on and off of low mass halos. When more halos are turning off, there would be a relatively flat part in the ionized fraction curve. Interestingly, the 512_all simulation completes reionization slightly sooner than the 256_all simulation, this despite the fact that HMACHs have dominated the photon budget by then. This result can be understood as a simple consequence that reionization completion depends on the total number of ionizing photons, which is higher for all redshifts in the 512_all simulation as compared to the 256_all simulation (Fig. 7.2).

To complete our presentation of the resolution study results, we show in Fig. 7.4 side-by-side projections of the logarithm of the neutral hydrogen fraction through the 6.4 Mpc volume at redshifts $z = 8, 7$ and 6. The color bar is chosen to show highly ionized gas as white, and partially ionized gas as shades of red-brown. The superimposed blue dots is the instantaneous ionizing emissivity field. One can see the larger number of smaller H II regions at earlier redshifts in the higher resolution simulation, as compared to the lower resolution simulation. Many of these are relic H II regions as their sources have turned off according to our probabilistic model of star formation in low mass halos. Once can also see that reionization has progressed further by $z = 6$ in the high resolution simulation, and that the strong ionization front driven by sustained star formation in the upper right corner of the cube is sweeping over smaller active and relic H II regions from earlier star formation in smaller halos.
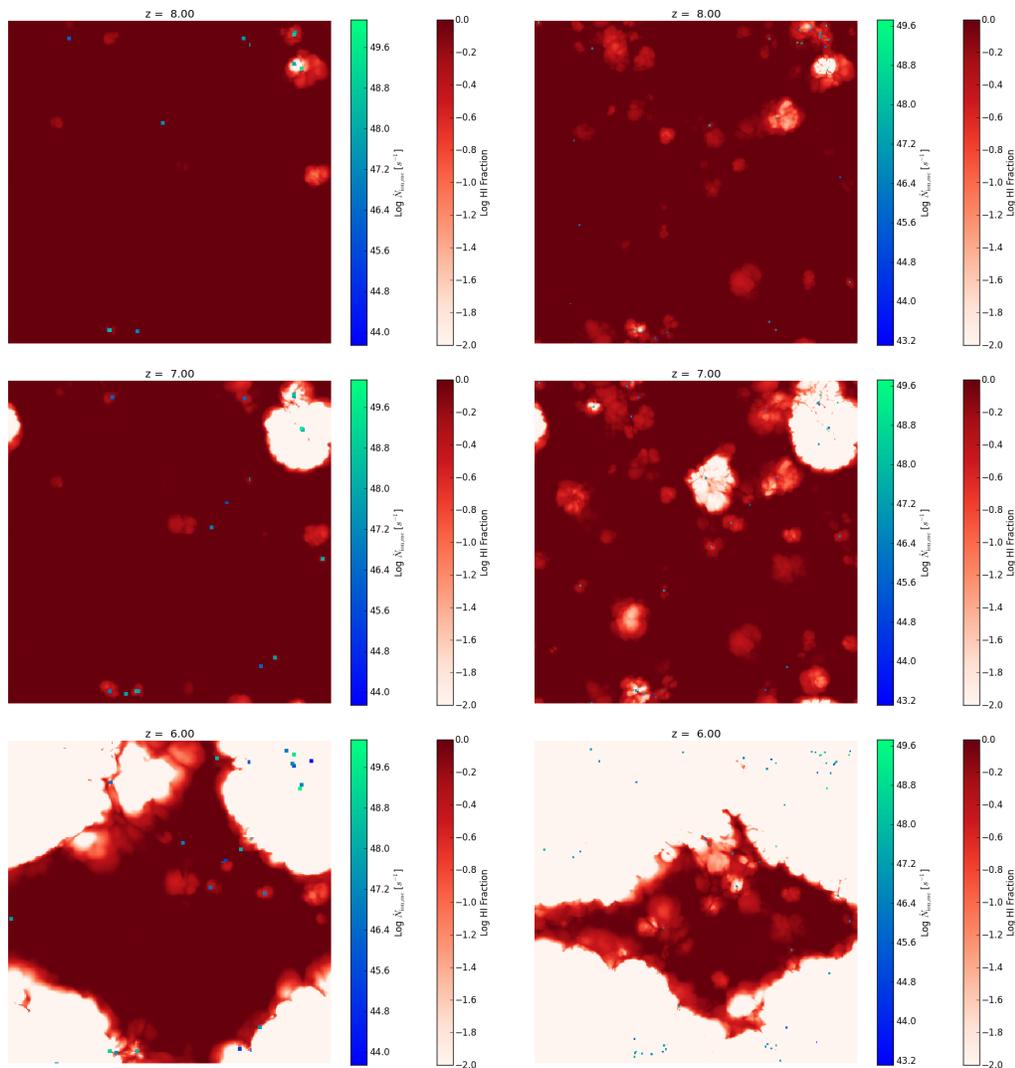
Figure 7.4: Projections of the logarithm of the neutral hydrogen fraction through the 6.4 Mpc volume at redshifts $z = 8, 7$ and 6 for the 256^3 (left column) and 512^3 (right column) test simulations. Note the increase in the number of relic H II regions in the high resolution simulation.

We explore this topic more thoroughly in the next section.

7.5.2   Science run–high resolution but a larger volume

Here we present the results of the science run carried out at identical mass and spatial resolution to the 512_all run, but in a box 2.25 times the size. Because the volume is greater than $10\times$ that of the former, we have much better statistical coverage of the ionizing sources at all redshifts.

Fig. 7.5 shows how reionization proceeds through a series of projections of the H I fraction through the box. The color table is chosen to accentuate the small H II regions of low to moderate ionization fraction, while larger highly ionized H II bubbles appear white. Superimposed as blue and green pixels is the ionizing emissivity field. One sees that before the HMACHs begin to dominate the total ionizing budget at $z \sim 8$, the volume is filled with small H II regions which are only partially ionized. They increase in size and number, but are still largely isolated at $z = 9$. By this time, a cluster of higher mass galaxies forms in the upper right hand corner of the box, and their combined ionizing flux drives a strong ionization front into the IGM. Because of our small volume and periodic boundary conditions, this H II superbubble fills the entire volume by $z = 7.1$, sweeping over the smaller H II regions as well as a smaller superbubble percolating in the center of the box.

Fig. 7.6 shows the redshift evolution of the number of ionizing photons escaping from halos in various mass ranges. The MCs (red line) begin contributing at $z \sim 22$ and dominate the LMACHs (green line) at all redshifts. This is due to their higher numbers and escape fractions as compared to the LMACHs. In fact the MCs dominate the HMACHs (turquoise line) until $z \sim 10$, and become subdominant thereafter. The
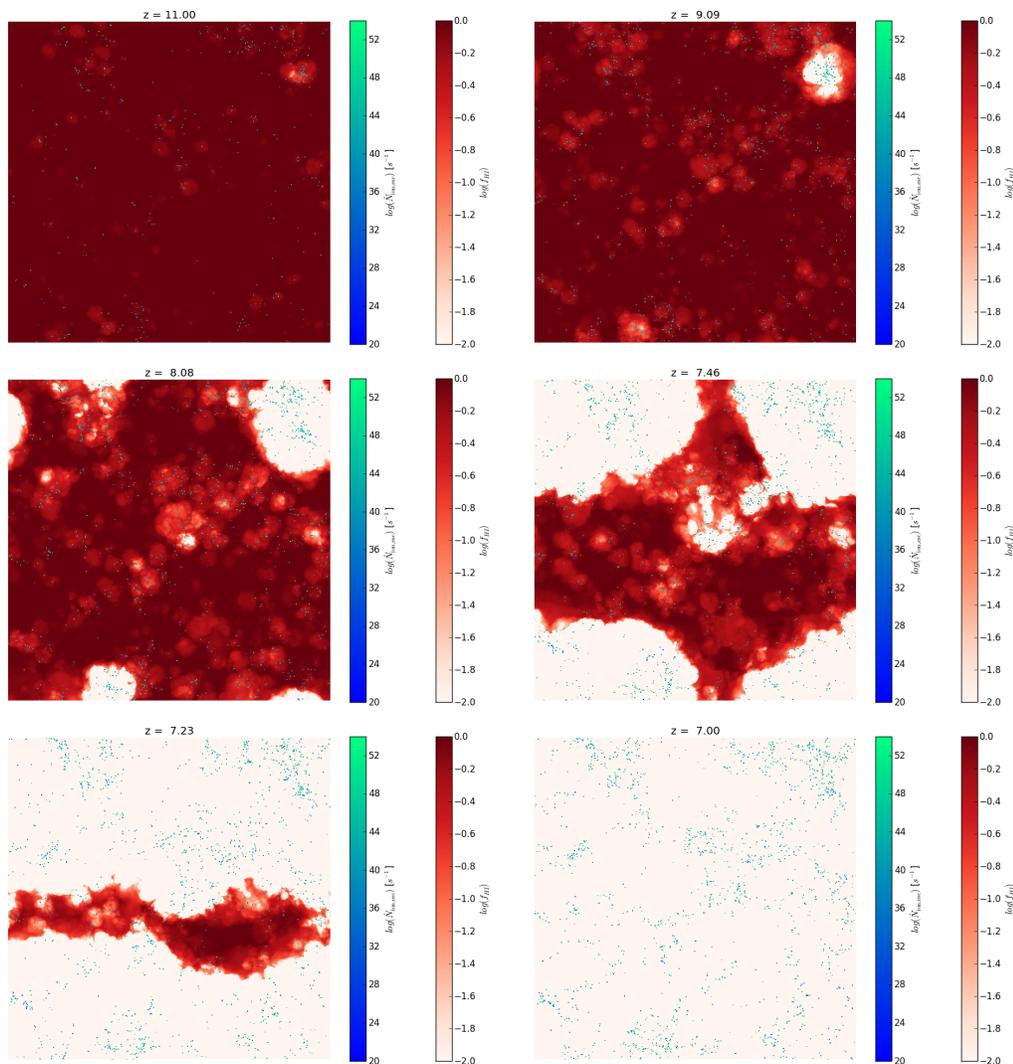
Figure 7.5: Projections of the logarithm of the neutral hydrogen fraction and ionizing emissivities (blue points) through the 14.4 Mpc volume at redshifts $z = 11, 10, 9, 8, 7.46$ and 7. This 1152^3 simulation has identical mass and spatial resolution as the 512^3 test simulation, but ionizes considerably earlier.
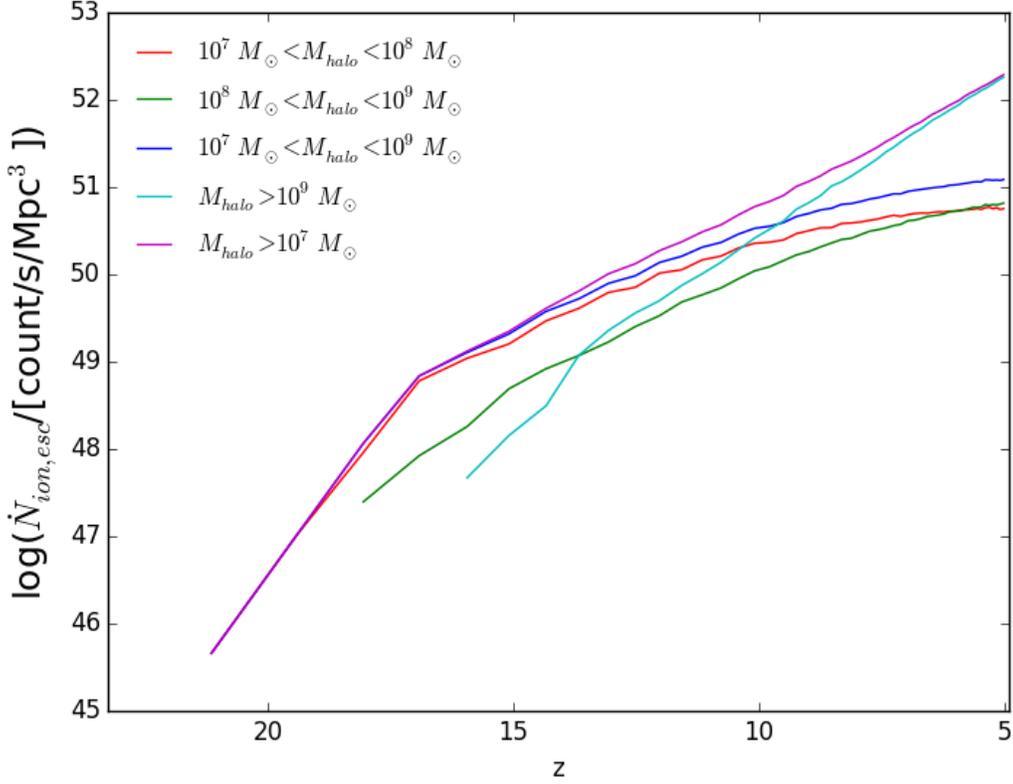
Figure 7.6: Number of escaping ionizing photons coming from halos in different mass bins from the 1152_all simulation. Blue line: $10^7 \leq M_{halo}/M_\odot \leq 10^8$; green line: $10^8 \leq M_{halo}/M_\odot \leq 10^9$; red line: $10^7 \leq M_{halo}/M_\odot \leq 10^9$; turquoise line: $M_{halo}/M_\odot > 10^9$. The purple line is the sum over all halos.

total ionizing photon flux is shown by the purple line, and increases by three orders of magnitude from $\dot{N} \sim 10^{49} \mathrm{s}^{-1}\mathrm{Mpc}^{-3}$ at $z = 17$ to $\dot{N} \sim 10^{52} \mathrm{s}^{-1}\mathrm{Mpc}^{-3}$ at $z = 7$, when overlap occurs.

The relative contribution of the different halo mass bins is illustrated in Fig. 7.7. As expected, MCs dominate the high redshift ionizing photon budget due to their high numbers and escape fractions. Interestingly, LMACHs are never more than a $\sim 20\%$ contributor, due to their significantly lower escape fractions. The HMACHs begin forming at $z \sim 16$ in this simulation, and only begin to exceed the contribution of the MCs at $z \sim 10$, and of MCs+LMACHs at $z \sim 8$. This figure makes it clear that
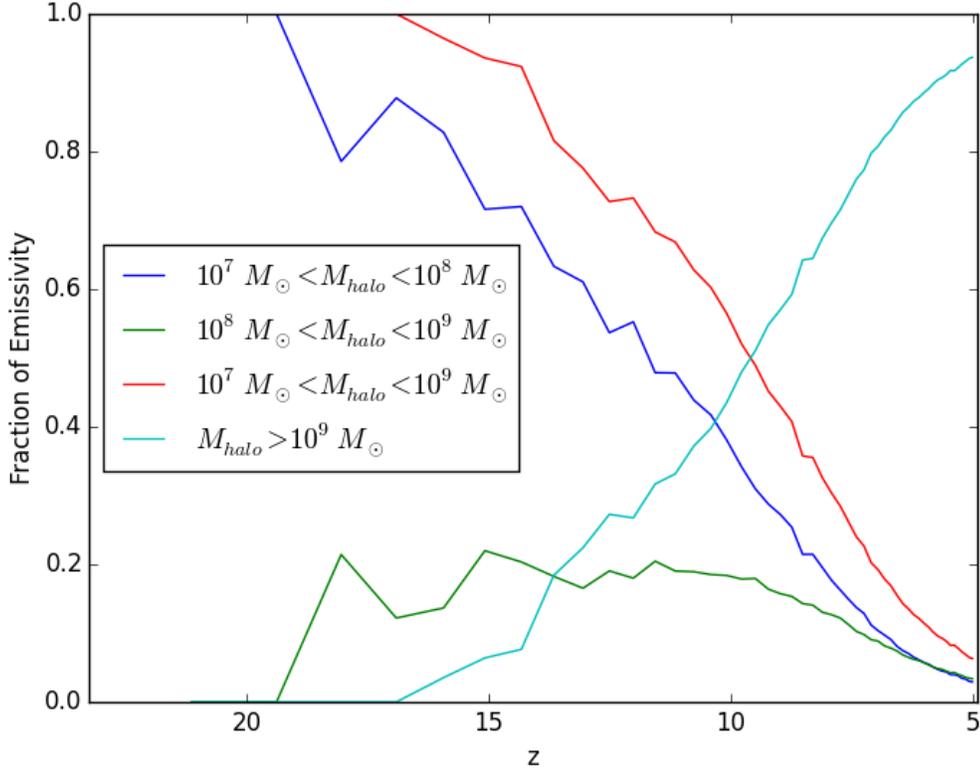
Figure 7.7: The fraction of the total ionizing luminosity coming from halos in different mass bins from the 1152_all simulation. Blue line: $10^7 \leq M_{halo}/M_{\odot} \leq 10^8$; green line: $10^8 \leq M_{halo}/M_{\odot} \leq 10^9$; red line: $10^7 \leq M_{halo}/M_{\odot} \leq 10^9$; and turquoise line: $M_{halo}/M_{\odot} > 10^9$.

the contribution of the MCs to the early phases of reionization $15 \geq z \geq 10$ cannot be ignored, and is more significant than that of the LMACHs.

Fig. 7.8 depicts the state of affairs at $z = 10$, when HMACHs begin to dominate the photon budget. Fig. 7.8a is a projection of the ionizing emissivity field. We see the volume is filled with many hundreds of ionizing sources. The sources are clustered; one can see in Fig. 7.8b the sources trace the density field.

Figs. 7.8c and 7.8d show slices of H I fraction and temperature through the most massive halo in the box at $z = 10$. One sees a couple of larger H II regions with
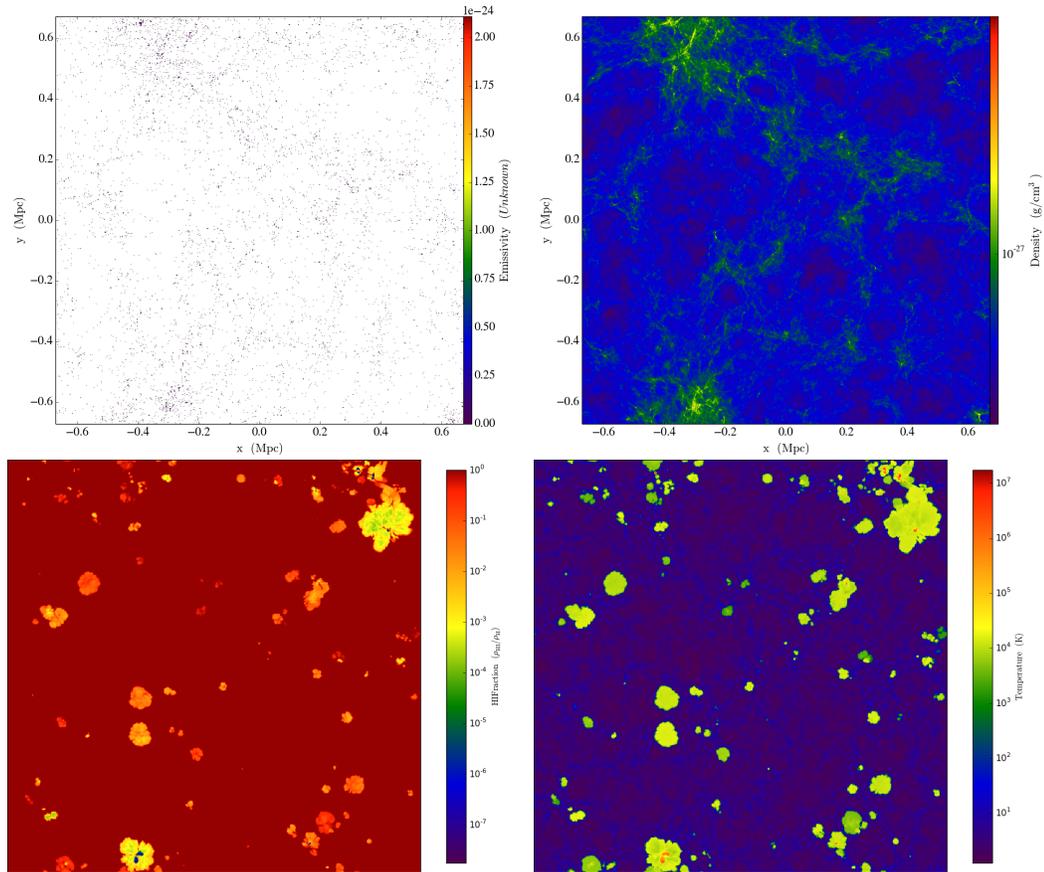
Figure 7.8: Top: projections of the logarithm of the baryon density and ionizing emissivity fields through the 14.4 Mpc volume at redshifts $z = 7.46$. Bottom: thin slices of the H I fraction and gas temperature fields at the same redshift.

$T > 10^4$K and $f_{HI} < 10^{-3}$ and numerous smaller H II regions of lower temperature and higher neutral fraction. The former mark the locations of HMACHs, which form stars continuously in our model, while the latter mark the locations of MCs and LMACHs, which form stars intermittently below $M_{halo} = 10^{8.5} M_\odot$, resulting in smaller, cooler, recombining H II regions

Fig. 7.9 shows the evolution of the ionized volume fraction f_i for three levels of ionization fraction (So et al., 2014): 10%, 99.9%, and 99.999%. The left(right) panel shows the linear(log) of the ionized volume fraction, respectively. Looking at the left panel first, we see that low levels of ionization (10%) are obtained in larger fractions of the volume than high levels of ionization $\geq 99.9\%$ at all redshifts, but is more pronounced at high redshifts. As found by So et al. (2014), the curve for the highest level of ionization 99.999% is significantly displaced to lower redshifts relative to the other two, and reaches $f_i = 1$ at $z = 6$, a $\Delta z = 1$ later.

Looking at the right panel, we see by the blue curve that lower levels of ionization begin to occupy tiny fractions of the volume before $z = 20$, consistent with the photon production history shown in Fig. 7.6. The blue curve increases monotonically to lower redshifts, reaching $f_i = 1$ at $z = 7.1$. The green curve shows the fraction of the volume that reaches the threshold of 99.9% local ionization fraction. It is not monotonic, but shows a sawtooth like modulation. This is a consequence of our insertion of a new set of ionizing sources every 20 Myr. While the periodicity is an artifact of our insertion method, some variability in f_i would be expected at early times in the continuous insertion limit as star formation in low mass halos turns on and off, creating relic H II regions in the process. This is particularly evident in the animated version of Fig. 7.5.
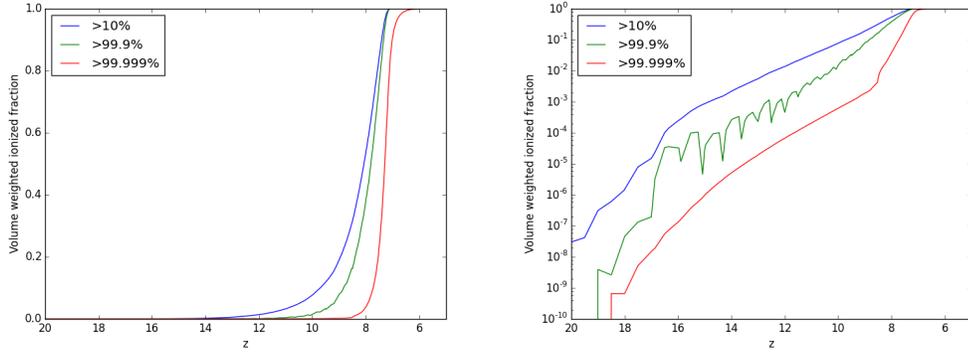
Figure 7.9: Ionized volume fraction versus redshift for different ionization fraction thresholds. Top: linear scale, Bottom: logarithmic scale. Relic H II regions show up prominently in the 99.9% curve as sources are reintroduced on 20 Myr intervals.

7.6   Discussion and Conclusions

In this paper, we study the role of low mass halos in reionization. After comparing two simulations with/without emissivities from smallest galaxies($10^7 M_\odot <$ $M_{\mathrm{halo}} < 10^9 M_\odot$) we conclude that the smallest galaxies boost τ_{es} and change the morphology of H II regions, but have little effect on the end of the reionization, which is dominated by HMACHs. We did two more simulations to study the effect of spatial and mass resolution in reionization. With higher spacial resolution and thus more smallest galaxies, the reionization begins earlier and the τ_{es} is larger, but the redshift when the reionization completes only advances a little($\Delta z \sim 0.2$).

The main weakness of our results is due to the limited resolution. Halo mass functions at every redshift are lower than the warren fit. As a result, we can't give a reliable prediction of the start of reionization with our simulations here.

Another weakness is that our simulation volume is not large enough to get the reliable characteristics of reionization. According to Iliev et al. (2014), a comoving

volume of $\sim$100 Mpc/h per side is sufficient for simulating a convergent mean reionization history. In the future we will do a fully coupled simulation with such volume size to study the reionization history and other reionization observables, including the impact on the Lyα forest.

The τ_{es} we calculate is not reliable, partially due to the cosmological model we use(WMAP 7) is not up to date, but mainly due to the mismatch in HMF. We don't generate enough halos so the τ_{es} we get is underestimated.

Chapter 7, in full, is currently being prepared for submission for publication with the title "Fully Coupled Simulation of Cosmic Reionization. III. Contribution of the Smallest Galaxies". Authors are Chen, Pengfei; Norman, Michael; Xu, Hao; Wise, John. The dissertation author was the primary investigator and author of this material.

Chapter 8

Conclusions and Open Questions

In this final chapter we summarize our main results and discuss open questions and improvements on the current work.

8.1   Main results

In this work we model quasars as radiating point sources to replace the homogeneous, evolving He II ionizing background in standard LAF simulations, and analyze thermal histories and the spectral properties of such models. Our main results are as follows:

1) Treating quasars as point sources, and using multigroup flux-limited diffusion to transfer radiation in multiple energy groups, we get better agreement with observations regarding the thermal history of the IGM at $z < 4$ compared with standard simulations using a UV background as the only ionizing source. The agreement lies both in the IGM temperature at mean density, and in the Doppler parameter b_σ from observed versus simulated H I LAF spectra (rescaled). In particular, we find that

inhomogeneous He II reionization results in a later peak ($z \sim 2.7$ for 1024_80_z5 and 1024_80_z7, and $z \sim 3.0$ for 1024_126_z5 and 1024_126_z7, $z \sim 3.6$ for 1024_80_BG) in the IGM temperature at the mean density, and it raises the temperature by several thousand K, which is consistent with recent findings in La Plante et al. (2016). At $z < 3.5$ b_σ's in inhomogeneous He II reionization simulations are higher than homogeneous optically thin simulation by ~ 1 km/s, which agrees with Paschos et al. (2007).

2) We see large sightline-to-sightline variation in the He II effective optical depths at $z > 2.5$ due to patchy He II reionization, with a scatter which is consistent with the observations of Worseck et al. (2014), and Fechner et al. (2006). The end of He II reionization in our large box simulations is $z \sim 2.76$, which agrees with Worseck et al. (2011), and Shull et al. (2010).

3) Even though inhomogeneous He II reionization changes the thermal history of the LAF, it doesn't have much effect on its spectral properties, which is consistent with results in La Plante et al. (2016) who performed simulations similar to our own. At $z \sim 2$, our H I LAF flux PDF agrees with data well in the range $0.3 < F < 0.8$, but shows departures in the complementary ranges. In particular, we have a lower probability of $F \sim 0$ flux than observations at $z \sim 2.0$, 2.5, and 3.0. With quasars as point sources, we get a lower amplitude 1D power spectrum, and it is still lower than observations over most of the range. For example, at $k \sim 10^{-2}$s/km we are missing $\sim 50\%$ in the amplitude comparing with observed data.

4) Including quasars in the redshift range $5 < z < 6.5$ has little effect on both the thermal history and spectral properties of LAF at $z < 4$. This shows that the IGM retains only a short-term memory of reionization events, and the effects from

the high redshift ionizing sources are averaged out by later photoheating and cosmic expansion (Hui and Haiman, 2003).

Our principal numerical innovations consist of the introduction of two methods to place ionizing sources using Enzo and yt alternately during a running simulation (or a job submission). Specifically:

1) To include galaxies at high redshifts as point sources, we use inline Python to find halos and measure their masses to find the right ionizing strength of their hosted galaxies, and modify emissivity fields accordingly. With inline Python yt could be called without stopping and restarting Enzo. See Appendix A.1 for details.

2) To include quasars at lower redshifts as point sources, every 45 Myr we stop Enzo, do halo finding with yt outside of Enzo, and then calculate the current radiation energies in each energy group for each quasar, and use them to replace old quasars. See Appendix A.2 for details.

8.2   Open questions and future plans

1) According to the resolution study in Lukić et al. (2015), we need a grid resolution of 20 kpc/h to produce 1 per cent convergence of H I LAF flux statistics. When including quasars as point sources, our current best grid resolution is ∼80 kpc/h, so we need to conduct simulations with 4 times higher resolution. At this resolution, a 80 Mpc/h box will require 4096^3 cells, quite an expensive simulation. A higher spatial resolution (a) will better model the high density regions, which may result in an increase in the flux PDF at $F \sim 0$ (See Fig. 5.7 for an evidence) and an increase in flux power spectrum and (b) will reduce the Hubble broadening of each cell length, and thus will generate a more reliable estimate of b_σ, especially at high redshifts. Of

course, even higher spatial resolution will be required to resolve galaxies and the H I halos surrounding them, which may be important to matching the observed flux power spectrum.

2) In this work we only replace the He II ionizing and heating background by discrete quasar sources with 5 radiation energy groups above 4 Ryd, while the H I and He I reionization is still done by the homogeneous UV background. In the future we can add radiation energy groups with energy between 1 to 4 Ryd. Specifically, (a) we can add contributions from quasars using the current or updated quasar luminosity function, and (b) we can add galaxies as point sources from high redshift, using the similar methods described in Chapter 7. MGFLD will still be used for radiation transfer. It may benefit us in at least two ways. (a) We can get a high redshift thermal history that is more consistent with observations. Evidence of improved agreement can be seen from Fig. 4.7, where simulation 1152, with galaxies as point sources, produces higher IGM temperature at $z > 5$ than simulation 1024_80_z7, and tends to better agree with observed data at $z \sim 5$. (b) With more H I-ionizing point sources, we expect larger fluctuations in H I density fields, which might increase the power of H I LAF on intermediate scales.

3) In this work we neglect the contribution from quasars that are too rare to be included in the box. However, they are very luminous and might contribute significantly to He II reionization. One approach for including them is to calculate the missing quasars' contribution according to the luminosity function and include that flux in the background. This might change both the thermal history and the spectral properties of the LAF.

4) Currently we don't have enough massive halos in our simulation box to

satisfy the halo mass-quasar luminosity one-to-one mapping relation (like footnote 14 in McQuinn et al. (2009)) and the quasar luminosity function at the same time. So in our simulations we loop over halos and assign radiation energy to them to satisfy the luminosity function only. But in this way (a) we may not match the clustering effects of halos in nature, and (b) we always pick same large halos to put quasars, even though their luminosity changes every 45 Myr. The way to improve might involve changing halo finding parameters, or changing halo finding methods, or try other methods to populate halos with quasars (La Plante et al., 2016).

5) To calculate radiation energy densities of several energy groups of a quasar given its bolometric luminosity, we need an assumption of its spectral energy distribution (SED) and also a relationship to convert the bolometric luminosity into specific luminosity. We currently use the code given by Hopkins et al. (2006) (see footnote 1 at Section 4.1). However, as the author mentions, the code is poorly adapted to handle photons with energies 1 to 10 Rydberg. So later on we could try other conversion relationships (e.g. Richards and others (2006)) and use latest quasar luminosity functions (e.g. McGreer and others (2013), Masters and others (2012), and Ross and others (2013)).

Appendix A

Software

Here we document the code we use to set up H I / He II reionization simulations, and the code to perform the spectral analysis of them. Unless otherwise specified, all scripts can be downloaded from `https://bitbucket.org/mlnorman/chen-thesis/`.

A.1   H I reionization simulations setup

For standard reionization simulations using a homogeneous UV background only, we just choose one background by setting RadiationFieldType in Enzo parameter file. To add additional point ionizing sources, the current way is to turn on star maker in Enzo. But we do not have enough flexibility in setting the locations and individual emissivities by doing that. Here we describe a new method to use inline python to add ionizing sources, which is the method we used in Chapter 8. With this method we can assign emissivities to halos while the Enzo job is running, according to the inline halo finding results and the halo mass-emissivity relation we have defined. This relation can be described in either a formula or a data table, and can be updated externally

without changing the Enzo source code. This offers a great deal of flexibility, and allows us to incorporate results from higher resolution simulations on properties such as the ionizing escape fraction.

A.1.1 Preparation

To get Enzo to work with inline yt, we need to install yt first. We use yt-2.x here. To download the install script, do

curl -O https://bitbucket.org/yt_analysis/yt/raw/yt-2.x/doc/install_script.sh

then in *install_script.sh* change the DEST_DIR, and change

./configure –prefix=$DEST_DIR/ $PYCONF_ARGS

into

./configure –enable-shared –prefix=$DEST_DIR/ $PYCONF_ARGS

, then install yt.

Then we move on to install Enzo. Here we use the stable branch of Enzo, obtained from `http://enzo-project.org`. We first activate yt that was just installed, then install hdf5 and hypre and set its path at LOCAL_HDF5_INSTALL and LOCAL_HYPRE_INSTALL respectively in the relevant makefile, i.e. *Make.mach.[machine _we_use]*. Also change

LOCAL_INCLUDES_HYPRE = -I$(LOCAL_HYPRE_INSTALL)/include

and

LOCAL_LIBS_HYPRE = -L$(LOCAL_HYPRE_INSTALL)/lib -lHYPRE

in *Make.mach.[machine_we_use]*.

In order to have a better control over inline python, we changed Enzo source files to add two more parameters. The files we change are saved under *enzo_change*.

Instead of moving those files to the source code folder of Enzo to replace the old ones, please use command "diff [old_file] [new_file]" to see the changes we made relevant to inline python and add/change them in the current file.

Now we can install Enzo. When we compile Enzo we set

make photon-yes

make hypre-yes

make emissivity-yes

make opt-high

in the config file in order to make FLD and inline python work properly.

A.1.2   Initial conditions

The files used to initiate simulation 1152_10 are saved under *chen-thesis/1152_i-nline*. Among them,

run1.inits is the input parameter file of *inits.exe* and it should be run on a large-memory node with at least 1 TB of RAM;

run1.enzo contains most parameters required by Enzo. The ones essential for the FLD solver are

```
RadHydroParamfile = RHParameters      // RHD module input parameters
StarMakerEmissivityField = 1 // use Geoffrey_s emissivity generator
RadiationFieldType = 0 // UV background (only usable if MultiSpecies=1)
ImplicitProblem = 3   // use the gFLDSplit module (Split Solver)
RadiativeTransfer   = 0
RadiativeTransferOpticallyThinH2 = 0
RadiativeTransferFLD = 2
RadiativeTransferHydrogenOnly = 0
```

The meaning of these parameters can be found in Enzo's user documentation here: `http://enzo.readthedocs.io/en/latest/parameters/index.html`. Parameters used to control inline python are

```
PythonTopGridSkip = 0
PythonSubcycleSkip = 0
PythonReloadScript = 1
PythonTimeSkip = 20 // Myr
```

Here PythonTimeSkip sets the time interval that we want Enzo to stop and run the inline python code, and it's in the unit of Myr;

RHParameters contains parameters related with RadHydro. Parameters that need more attention are RadHydroDtRadFac and RadHydroSolTolerance.

A.1.3 Inline python

What we do with inline python is realized in *user_script_run1.py*, and it's shown below. It modifies the emissivity field inline every PythonTimeSkip Myr. Here "inline" means doing it without stopping Enzo. *table_new.txt* contains emissivity versus halo mass information derived from the Renaissance Simulations with sub-kpc resolutions (Xu et al., 2016b) and it's saved under *chen-thesis/1152_inline*. It is a 20×50 table, which is the probability distribution function of escaped ionizing photons per second $\dot{N}_{ionizing,esc}$, given halo mass M_{halo}. The table is reproduced in Chapter 8, Table 1. The range of halo mass applicable to this table is 10^7 to 10^9 $M_\odot$. So 20 rows in the table correspond to halos with mass in the range of $10^{7.0}$-$10^{7.1}$ $M_\odot$, ..., $10^{8.9}$-$10^{9.0}$ $M_\odot$, respectively. Given one row, we have 50 columns, which correspond to possible values of emissivity for that halo mass bin. Each of the emissivity has

probability 0.02 of occurring. After the halo finding we get the halo mass of each halo, and find its mass interval in the table, i.e the row of the table, then we draw from an uniform distribution and use that random number to choose the emissivity we use, i.e. the column in that row. Both *user_script_run1.py* and *table_new.txt* should be directly under the simulation folder. Before we start the simulation we rename *user_script_run1.py* to *user_script.py*, since that's the only name of inline python script that Enzo anticipates.

```python
# user_script_run1.py
# Purpose: to clear and re-assign emissivity fields
# Should change to user_script.py before starting the run.
# Make sure table_new.txt is also in this directory.
# It is called by the enzo source code CallPython.C.


from yt.mods import *
from yt.analysis_modules.halo_finding.api import *
from yt.utilities.parallel_tools.parallel_analysis_interface import \
    communication_system
import enzo
import numpy as na
import time
import random


field = 'Emissivity'
gz = 3 # size of the ghost zone
result_path = 'run1/results/'


def _HIFraction(field, data):
    return data["HI_Density"]/(data["HI_Density"]+data["HII_Density"])
```

```python
add_field("HIFraction", function=_HIFraction, units=r"\rho_\mathrm{HI}/\
    rho_\mathrm{H}")


# Convenient func to print to the log file
def pt(fn, st):
    fo = open(result_path+fn, "a")
    fo.write(st)
    fo.close()


# Func to clear the current emissivity field
def clear_emis():
    shape = na.array(enzo.grid_data[enzo.grid_data.keys()[0]][field]).
    shape
    for i in range(shape[0]):
            for j in range(shape[1]):
                    for k in range(shape[2]):
                            enzo.grid_data[enzo.grid_data.keys()[0]][field]\
                                [i][j][k]=0.0


# Func to check if the emissivity field is cleared, use if you need
def check_emis():
    a = na.array(enzo.grid_data[enzo.grid_data.keys()[0]][field])
    if a.max():
            pt("output_run1.txt","Error! Maximum %e\n"%a.max())


# Load the table containing PDF of \dot{N}_{ionizing, esc}
# versus M_{halo}
table = na.loadtxt("table_new.txt")
```

```python
dm = 0.1 # interval in log mass

m0 = 7.0 # minimum log10(M_halo) to use the table

m1 = 9.0 # maximum log10(M_halo) to use the table

np = 50  # No. of bins in probability


# Func to return the emissivity (in erg/s) given the halo mass
# Assume that the average energy of emitted photons is 21.6eV
def table_emis(m):
    if na.log10(m) >= m1:
        return na.power(10,36.033+1.675*na.log10(m))*3.46070138e-11
                                                           # erg/s
    i0 = na.int((na.log10(m)-m0)/dm)
    # Use a uniform random number to choose the \dot{N}_{ionizing, esc}
    return table[i0][random.randint(0,np-1)]*3.46070138e-11 # erg/s


def save(pf, plt, add=""):
    plt.save(result_path+"%s" % pf +add)


# Fields that will be plotted
# p for projections
# s for slices
p_fields = ['HIFraction', 'Density', 'Emissivity', 'Temperature']
s_fields = p_fields


# Func to make projections and slices
def proj_slice(pf, halos, label):
    for field in p_fields:
        prj = ProjectionPlot(pf, 2, field,
                                weight_field = 'CellVolume')
```

```python
        # Add circles representing halos in projections
        prj.annotate_hop_circles(halos)
        only_on_root(save, pf, prj, label)
    prj = ProjectionPlot(pf, 2, 'Emissivity',
                         weight_field = 'Emissivity')
    prj.annotate_hop_circles(halos)
    only_on_root(save, pf, prj, label)
    for field in s_fields:
        pc = PlotCollection(pf)
        # Need to specify the center!
        pc.add_slice(field, 2, center=[0.5, 0.5, 0.5])
        only_on_root(save, pf, pc, label)


def main():
    pf = EnzoStaticOutputInMemory()
    z = pf.current_redshift
    t = pf.time_units['Myr']*pf.current_time
    # Write info into the log file
    only_on_root(pt,"output_run1.txt","%s z=%23.16e t=%23.16e %s\n" \
                 %(pf,z,t,time.strftime("%Y-%m-%d %H:%M:%S",
                                        time.localtime()))))
    comm = communication_system.communicators[-1]
    c2 = "%s"%enzo.grid_data.keys()
    c2 = na.int(c2[1:-2]) # ID of current core / task
    cw = 1.0/1152.0 # cell width in code units
    # Do the following calculations, i.e. placing ionizing sources,
    # only after z~30
    if z<30:
        # Halo finding
```

```python
        halo_list = parallelHF(pf, dm_only=True, threshold=80.0)
        halo_list.dump("run1/HaloDir/InlineHalos_%.2f"%z)
        # Make plots
        proj_slice(pf, halo_list, "_%.2f"%z)
        # Do the following only if we have halos > 10^7 M_{sun}
        if na.alen(halo_list) and halo_list[0].total_mass()>1e7:
            # Save emissivities from halos with mass satisfying
            # 7<log10(M_{halo}/M_{sun})<8 , etc.
            emis_78 = 0.0
            emis_89 = 0.0
            emis_9 = 0.0
            emis = 0.0
            clear_emis()
#            check_emis()
            # Loop over qualified halos to change emissivity fields
            for halo in halo_list:
                mass = halo.total_mass()
                if mass>1e7:
                    if comm.rank == 0:
                        emis = table_emis(mass)/(cw**3*pf.units['cm']**3)
                                                            # erg/s/cm^3
                        if mass>1e9:
                            emis_9 += emis
                        elif mass>1e8:
                            emis_89 += emis
                        else:
                            emis_78 += emis
                        emis /= 27.0
                    emis = comm.mpi_bcast(emis)
```

```
                    # Spread the emissivity equally into 3x3x3 cells,
                    # centered in that halo
                    for i in [-1, 0, 1]:
                        for j in [-1, 0, 1]:
                            for k in [-1, 0, 1]:
                                coord = (na.array(halo.center_of_mass()) +\
                                        na.array([cw*i, cw*j, cw*k]))%1
                                this = pf.h.find_point(coord)[0][-1]
                                c1 = "%s"%this
                                # ID of the core in charge of that cell
                                c1 = na.int(c1[-4:])
                                if c1==c2:
                                    mark = na.zeros(3).astype('int')
                                    # Calculate the index of that cell
                                    for dim in xrange(len(coord)):
                                        mark[dim] = int((coord[dim] - \
                                                    this.LeftEdge[dim]) / cw)
                                    # Add emissivity into the current one
                                    # instead of replacing it
                                    enzo.grid_data[enzo.grid_data.keys()[0]]\
                                        [field][gz+mark[2]][gz+mark[1]]\
                                        [gz+mark[0]] = enzo.grid_data\
                                        [enzo.grid_data.keys()[0]][field]\
                                        [gz+mark[2]][gz+mark[1]][gz+mark[0]]+\
                                        emis
            else:
                break
    del halo_list
    # Save the statistics of emissivities into the file
```

```
        only_on_root(pt,"Emis_run1.txt","%23.16e %23.16e %23.16e
    %23.16e %23.16e\n" \
                            %(z, emis_78, emis_89, emis_9, emis_78+
    emis_89+emis_9))
```

A.2   He II reionization simulations setup

A.2.1   Initial conditions

To use MGFLD for radiation transfer, we use Dr. Daniel R. Reynolds's version of Enzo, which could be downloaded from `https://bitbucket.org/drreynolds/enzo-dev_reynolds`. When we compile Enzo, we turn on photon, hypre, emissivity, and turn off python. The reason we turn python off is that we do not perform inline halo finding in this simulation; halo finding is done between restarts for reasons described below. The files used to initiate simulation 1024_80_z7 are saved under *chen-thesis/1024_80_z7*. Among them,

1024_80Mpc_z7.inits is the input parameter file of *inits.exe* and in this case it should be run with a large-memory node due to the large number of grids;

1024_80Mpc_z7.enzo contains most parameters required by Enzo. The ones essential for MGFLD are

```
RadHydroParamfile = 1024_80Mpc_z7.amrfld // RHD module input parameters
StarMakerEmissivityField = 1 // use Geoffrey_s emissivity generator
ImplicitProblem   = 6        // use AMRFLDSplit module
RadiativeTransfer  = 0
RadiativeTransferOpticallyThinH2 = 0
RadiativeTransferFLD = 2
```

We also specify redshifts of all data dumps we need in this file. To calculate the redshifts 45 Myr apart given a certain set of cosmology parameters, we use code *time_to_z.py* saved in *chen-thesis/gadget*;

1024_80Mpc_z7.amrfld contains parameters related with the MGFLD solver. Initially we set 5 frequency bands (which may be changed) but with no sources, so that 5 radiation energy fields and 5 emissivity fields are created from the beginning of the simulation. Multiple tests have shown that

AMRFLDSolTolerance $\quad = 1.0\,\mathrm{e}-10$

is the maximum tolerance to guarantee accuracy in the radiation transfer;

hm12_photorates.dat is the table containing the photo-ionization and photo-heating rates of H I, He I and He II from Haardt and Madau (2012) (HM2012). Here we remove the rates related with He II photo-ionization and -heating at redshifts $z < z_{\mathrm{QSO}}$ ($z_{\mathrm{QSO}}=6.5$ in this case), since we will explicitly include quasars in the volume to account for those He II ionizing sources. The script to do this modification is *change_hm12.py* and is saved under *chen-thesis/gadget*. For the original version of the table, see *enzo/input/hm12_photorates.dat* in the Enzo repository;

run1.pbs is the submission script that will run the simulation down to redshift z_{QSO} when we will begin placing quasars.

A.2.2 Placement of quasars

After we get to z_{QSO}, we put quasars into our simulation box according to the observed quasar luminosity functions. See Section 4.1 for more descriptions. Here we don't use inline python, since now we are using MGFLD for radiation transfer, and, according to our tests, we cannot use inline python to change the parameters that

MGFLD uses to define the properties of ionizing sources. So every 45 Myr we stop the simulation, update the placement of quasars, and restart the simulation, which is done by *run2.pbs*. Its essential part is

```
sim=1024_80Mpc_z7
dir=/scratch/sciteam/madcpf/$sim/Dumps
source /u/sciteam/madcpf/yt-x86_64-shared/bin/activate
n="0"
while [ $n -lt 71 ] # 71 is the max id of RD dumps in this case
do
full_file=`ls $dir/?D????/?D???? -t | head -1`
file=`echo "${full_file:(-6)}"`
n=`echo "${file:(-4)}"`
if [[ $file == RD* ]]
then
python check_restart_MGFLD_all.py $sim
aprun -cc 0,2,4,6,8,10,12,14,16,18,20,22,24,26,28,30 -n 1024 \
python ParallelHF_all.py $sim $file --parallel
python MGFLD_allLF_with_highz_all.py $sim $file
fi
aprun -cc 0,2,4,6,8,10,12,14,16,18,20,22,24,26,28,30 -n 1024 \
./enzo071116.exe -d -r $full_file 2>&1 | tee \
restart_`printf %s $sim`_$file.txt
done
```

Every time the simulation job runs out of time, we just submit the same submission script and it will find the latest data dump, do the halo finding using yt, do calculations according to the halo finding results and the observed quasar luminosity functions at that redshift, write the results into the parameter file and log file, and

restart from it. Here we only do this if the data dump starts with "RD", so that if we add more outputs with "DD" outputs according to number of cycles, they won't affect the results. More specifically, those are done by the following three scripts. Note that in these codes we use yt-2.x.

First we use the following code to locate the latest data dump and change its redshift parameters for the next restart.

```python
# check_restart_MGFLD_all.py
# Purpose: to modify parameter CosmologyFinalRedshift
# in file RD????/RD????
# It is called by run2.pbs.
# To run it:
# python check_restart_MGFLD_all.py <simulation>


import os
import numpy as na
import time


Dumps_add = ''
sim = sys.argv[1]
data_path = "/scratch/sciteam/madcpf/%s/Dumps%s/"%(sim, Dumps_add)
result_path = '/scratch/sciteam/madcpf/%s/'%sim


# Find the latest data dump
latest = max([os.path.join(data_path,d) for d in os.listdir(data_path)],
             key=os.path.getmtime)
fn = latest+'/'+latest.split('/')[-1]
print "check_restart.py"+fn
# Rename the original parameter file
```

```python
command = 'mv %s %s_origin '%(fn, fn)
os.system(command)


fin = open(fn+'_origin')
lines = fin.readlines()
fout = open(fn, "w")
z_final = 2.0
found = False


# Find the redshift corresponding to 45 Myr later
# and set it to be CosmologyFinalRedshift
for line in lines:
    if line.startswith('CosmologyCurrentRedshift'):
        z_now = na.float(line.split('=')[-1])
    elif (not found) and line.startswith('CosmologyOutputRedshift'):
        z_next = na.float(line.split('=')[-1])
        # Skip the z = 2.5, z = 3.0, z = 3.99 ... output,
        # which are added manually but not at 45 Myr interval
        if z_next==z_final or z_next*100>na.int(z_next*100):
            l_new = 'CosmologyFinalRedshift = %f\n'%z_next
            found = True
    elif line.startswith('CosmologyFinalRedshift'):
        line = l_new
    fout.write(line)
fin.close()
fout.close()


# Write info into the log file
fo = open(result_path+"log_restart%s.txt"%Dumps_add, "a")
```

```
fo.write("\n" +"#"*20+"\nrestart from %s, z=%f, %s\n"%(fn.split('/')\
    [-1], z_now, time.strftime("%Y-%m-%d %H:%M:%S",time.localtime())))
fo.close()
```

Then we use the code below to find halos if necessary.

```python
# ParallelHF_all.py
# Purpose: to find halos
# It is called by run2.pbs.
# To run it on Blue Waters:
# aprun -cc 0,2,4,6,8,10,12,14,16,18,20,22,24,26,28,30 -n 1024
# python ParallelHF_all.py <simulation> <file_id> --parallel


from yt.mods import *
from yt.analysis_modules.halo_finding.api import *
import sys
import numpy as na


Dumps_add = ''
sim = sys.argv[1]
file = sys.argv[2]
path = '/scratch/sciteam/madcpf/$s/'%sim
fn_add = sim
pf = load(path+'Dumps%s/%s/%s' % (Dumps_add, file, file))
res = pf.domain_dimensions[0] # grids of the simulation box
z = pf["CosmologyCurrentRedshift"]


# Skip the halo finding If it's not the 45-Myr redshift
if z<6.5 and z*100==na.int(z*100):
    sys.exit()
```

```
# Do halo finding and save the results
# A higher threshold results in less halos, which is bad
halo_list = parallelHF(pf, threshold = 80.0, dm_only = True,
                       num_particles = res**3)
halo_list.dump(path+"HaloDir%s/Halos_%s_%s_%f"%
               (Dumps_add, fn_add, file, z))
```

Then, with the halo finding results, we use the following code to calculate the number of quasars, the radiation energy in each energy bin for each quasar, and to use them to replace the old MGFLD parameters. At first we tried to use the mass-to-luminosity mapping function as in McQuinn et al. (2009), but then we found that we did not have enough large halos to satisfy the luminosity function. So we change to the current method where, instead of using one to one mapping, we start from the most massive halos and assume that they have the largest luminosity in the box, and so on. In the code we call an external program "qlf_calc", which is the quasar luminosity calculator from Hopkins et al. (2006). The source code is at http://www.cfa.harvard.edu/~phopkins/Site/qlf.html.

```
# MGFLD_allLF_with_highz_all.py
# Purpose: to calculate the number of quasars and radiation energy
# for each radiation energy bin for each quasar
# according to the observed quasar luminosity functions,
# and write them into parameter files together with the
# location of new quasars
# See Pengfei's thesis for more descriptions
# It is called by run2.pbs.
# To run it:
```

```python
# python MGFLD_allLF_with_highz_all.py <simulation> <file_id>
# --parallel


from yt.mods import *

import numpy as na

import os, sys

from yt.analysis_modules.halo_finding.api import *

import random

import time


# Constants

hplanck = 6.6260693e-27 # erg s

ev2erg = 1.60217653e-12

clight = 2.99792458e10 # cm/s

kboltz = 1.3806504e-16

nu0_HI = 13.6 / hplanck * ev2erg   # Hz

nu0_HeII = 54.4 / hplanck * ev2erg   # Hz

nu_1450 = clight/1450e-8 # Hz

L_sun = 3.9e33   # erg/s

M_sun = 4.76


alpha = 1.6 # from Hopkins 06

sigma_alpha = 0.2 # from Hopkins 06


sim = sys.argv[1]

file = sys.argv[2]

path = '/scratch/sciteam/madcpf/%s/'%sim

Dumps_add = ''

fn_add = sim
```

```python
pf = load(path+'Dumps%s/%s/%s' % (Dumps_add, file, file))
z = pf["CosmologyCurrentRedshift"]
# Skip the z = 2.5, z = 3.0, z = 3.99 ... output,
# which are added manually but not at 45 Myr interval
if z<6.5 and z*100==na.int(z*100):
    sys.exit()
L_box = pf["CosmologyComovingBoxSize"]/pf["CosmologyHubbleConstantNow"]
                                                # comoving Mpc
HaloListname = path+"HaloDir%s/Halos_%s_%s_%f"%(Dumps_add,fn_add,file,z)
halos = LoadHaloes(pf, HaloListname)
M_halo_max = halos[0].total_mass() # M_sun
L_Bol_max = na.power(10,15) # L_sun, from Hopkins06
L_Bol_min = na.power(10,10) # L_sun, from Hopkins06
L_1450_min = na.power(10, -0.4*(-18.5-M_sun))*L_sun/nu_1450 # erg/s/Hz
L_1450_max = na.power(10, -0.4*(-28-M_sun))*L_sun/nu_1450 # erg/s/Hz


os.system("qlf_calc %f %f 0 > %s/LF_with_highz_z%f.out"%
            (nu0_HI, z, path+"results",z))
nuLnu, Lbol = na.loadtxt("%s/LF_with_highz_z%f.out"%(path+"results",z),
                            usecols=(0, 3), unpack=True)


# LF Model from Eqn (2) in Giallongo 2015, used for 5<z<6.5
def dPhi_dlogL1450(z, log_l_1450):
    if z>4 and z<4.5:
        beta = 1.52
        gamma = 3.13
        M_break = -23.2
        phi_star = na.power(10, -5.2)
    elif z>4.5 and z<5:
```

```
        beta = 1.81

        gamma = 3.14

        M_break = -23.6

        phi_star = na.power(10, -5.7)

    elif z>5 and z<=6.5:

        beta = 1.66

        gamma = 3.35

        M_break = -23.4

        phi_star = na.power(10, -5.8)

    L_break =na.power(10, -0.4*(M_break-M_sun))*L_sun/nu_1450 # erg/s/Hz

    return  2.5*phi_star/(na.power(na.power(10,log_l_1450)/L_break,

            beta-1) + na.power(na.power(10,log_l_1450)/L_break, gamma-1)

    )

                                                # /comoving Mpc^3/logL_1450


# Func to calculate number of quasars per L bin when 5<z<6.5
def N_quasar_dlogL_1450(z,dlogL):

    lam=[]

    n = int((na.log10(L_1450_max)-na.log10(L_1450_min))/dlogL)

    bins = na.linspace(na.log10(L_1450_min), na.log10(L_1450_max), n+1)

    for i in range(n):

        xs = na.linspace(bins[i], bins[i+1],

                        int((bins[i+1]-bins[i])/0.01+1))

        intg = dPhi_dlogL1450(z, xs)

        lam.append(na.sum((intg[1:]+intg[:-1])/2.*0.01) * L_box**3)

    return n, na.array(lam), na.power(10,bins)


# LF Full model from Hopkins 06, used for z<5
P0=-4.8250643; P1=13.035753;    P2=0.63150872;   P3=-11.763560;
```

```python
P4=-14.249833;  P5=0.41698725;   P6=-0.62298947;  P7=2.1744386;
P8=1.4599393;   P9=-0.79280099;  P10=0.;           P11=0.;
P12=0.;          P13=0.;          P14=0.;
beta_min = 1.3
phi_star = P0
def dPhi_dlogL(z, log_l_bol):
    xsi = na.log10((1. + z)/(1. + 2.))
    l_star   = P1 + P2*xsi + P3*xsi*xsi + P4*xsi*xsi*xsi
    gamma_1  = P5 * na.power(10.,xsi*P6)
    gamma_2  = 2.0 * P7 / (na.power(10.,xsi*P8) + na.power(10.,xsi*P9))
    if gamma_2 < beta_min:
        gamma_2 = beta_min
    x = log_l_bol - l_star
    return na.power(10., phi_star - na.log10(na.power(10.,x*gamma_1) + \
                                              na.power(10.,x*gamma_2)))
                                                # /comoving Mpc^3/logL_bol


# Func to calculate number of quasars per L bin when z<5
def N_quasar_dlogL_bol(z,dlogL):
    lam=[]
    n = int((na.log10(L_Bol_max)-na.log10(L_Bol_min))/dlogL)
    bins = na.linspace(na.log10(L_Bol_min), na.log10(L_Bol_max), n+1)
    for i in range(n):
        xs = na.linspace(bins[i], bins[i+1],
                         int((bins[i+1]-bins[i])/0.01+1))
        intg = dPhi_dlogL(z, xs)
        lam.append(na.sum((intg[1:]+intg[:-1])/2.*0.01) * L_box**3)
    return n, na.array(lam), na.power(10,bins)
```

```python
# bin width for luminosity
dlogL_bol = 0.25
dlogL_1450 = 0.2533333333333333


# Calculate no. of L bins, no.of quasars of each L bin, L bins,
# with zero terms remained
if z<=5:
    N_L, N_q_lam, L_bols =  N_quasar_dlogL_bol(z,dlogL_bol)
else:
    N_L, N_q_lam, L_1450s =  N_quasar_dlogL_1450(z,dlogL_1450)


AMRFLDNumRadiationFields = 5
AMRFLDFrequencyBand = na.array([54.4, 65., 75., 125., 155., 400.]) # eV
nu = AMRFLDFrequencyBand/ hplanck * ev2erg # Hz
AMRFLDSourceLocation = []
AMRFLDSourceGroupEnergy = [] # erg/s


# Func to calculate the radiation energy in each energy bin
# using the integraion results of the power law spectrum
def calc_AMRFLDSourceGroupEnergy(L0):
    # each quasar has its own spectra (alpha_)
    alpha_ = na.random.normal(alpha, sigma_alpha)
    tmp = [L0*na.power(nu0_HI, alpha_)/(1.-alpha_) * \
                (na.power(nu[i+1], 1.-alpha_) - \
                    na.power(nu[i], 1.-alpha_))
            for i in range(len(nu)-1)] # erg/s
    AMRFLDSourceGroupEnergy.append(tmp)


# Write info into the log file
```

```python
f_log = open(path+"log_restart%s.txt"%Dumps_add, "a")
f_log.write("Halo finding and calc finished. %s\n"%
                time.strftime("%Y-%m-%d %H:%M:%S", time.localtime()))


# Calculate the radiation energy in each energy bin
# Use the center of mass of halos as the location of the quasar
N_halos = []
N_q = []
id_now = 0
total_halos = len(halos)
for i in range(N_L)[::-1]:  # start from the largest halos
    ni = na.random.poisson(lam=N_q_lam[i])
    id_end = min(id_now+ni, total_halos)
    for id in range(id_now,id_end):
        if z<=5:
            rand_Lbol = random.uniform(L_bols[i], L_bols[i+1])
            L912 = na.power(10,na.interp(na.log10(rand_Lbol*L_sun),
                                         Lbol, nuLnu))/nu0_HI
        else:
            rand_L1450 = random.uniform(L_1450s[i], L_1450s[i+1])
            # from Eqn (3) in Giallongo 2015
            L912 = rand_L1450*na.power(1200./1450., 0.44) * \
                na.power(912./1200., 1.57)
        calc_AMRFLDSourceGroupEnergy(L912)
        AMRFLDSourceLocation.append(na.array(halos\
                                    [id].center_of_mass())%1)
    N_q.append(id_end-id_now)
    id_now = id_end
    if id_end==total_halos:
```

```python
            f_log.write("not enough halos from lumnosity bin %d!\n"%i)

            break


# Write more info into the log file
N_q_sum = sum(N_q) # total number of quasars
N_q = N_q[::-1]
if z<=5:
    lstring = "L_bol"
    Ls = L_bols
else:
    lstring = "L_1450"
    Ls = L_1450s
f_log.write("M_halo_max               = %e M_sun\n"%M_halo_max +
            "Number of L_bol bins     = %d\n"%N_L +
            lstring + " bins               = "+str(Ls)+"\n"+
            "Number(lambda) of quasars = "+str(N_q_lam)+"\n"+
            "Number of halos chosen   = "+str(N_q)+"\n"+
            "Total lambda             = %d\n"%sum(N_q_lam)+
            "Total halos chosen       = %d\n\n"%sum(N_q))
f_log.write("AMRFLDNumSources = %d\n"%N_q_sum)
f_log.write("AMRFLDNumRadiationFields = %d\n"%AMRFLDNumRadiationFields)
for i in range(AMRFLDNumRadiationFields):
     f_log.write("AMRFLDFrequencyBand[%d] = %f %f\n"%(i,
                       AMRFLDFrequencyBand[i], AMRFLDFrequencyBand[i+1]))
for i in range(N_q_sum):
    f_log.write("AMRFLDSourceLocation[%d] = %f %f %f\n"%(i,
                  AMRFLDSourceLocation[i][0], AMRFLDSourceLocation[i][1],
                                        AMRFLDSourceLocation[i][2]))
    for j in range(AMRFLDNumRadiationFields):
```

```python
        f_log.write("AMRFLDSourceGroupEnergy[%d][%d] = %e\n"%(i, j,
                                        AMRFLDSourceGroupEnergy[i][j]))


# Rename the original MGFLD parameter file
# and write a new one
fn = path+'Dumps%s/%s/%s.rtmodule' % (Dumps_add, file, file)
command = 'mv %s %s_origin'%(fn, fn)
os.system(command)
f_in = open(fn+'_origin')
lines = f_in.readlines()
f_out = open(fn, "w")
for line in lines:
    if not line.startswith('AMRFLDNumSources') and \
            not line.startswith('AMRFLDSourceLocation') and \
            not line.startswith('AMRFLDSourceGroupEnergy'):
        f_out.write(line)
f_out.write("AMRFLDNumSources = %d\n"%N_q_sum)
for i in range(N_q_sum):
    f_out.write("AMRFLDSourceLocation[%d] = %f %f %f\n"%(i,
                AMRFLDSourceLocation[i][0], AMRFLDSourceLocation[i][1],
                                        AMRFLDSourceLocation[i][2]))
    for j in range(AMRFLDNumRadiationFields):
        f_out.write("AMRFLDSourceGroupEnergy[%d][%d] = %e\n"%(i, j,
                                        AMRFLDSourceGroupEnergy[i][j]))
f_in.close()
f_out.close()
f_log.close()
```

A.3 Spectral analysis

When a simulation is done, we choose several data dumps at redshifts of interest and do spectral analysis for each of them. We used both Blue Waters at NCSA and Comet at SDSC to do such analysis. The codes are different since we use Comet's SSD space to speed up the calculations. Here we first show our codes used on Blue Waters in detail, and then discuss the change in Comet briefly.

A.3.1 Blue Waters

Step 1: light rays - The first step is generating light rays, which samples H I / He II density, temperature, peculiar velocity fields along specific lines of sight for later analysis. yt light rays may pass through several data sets covering an interval of redshift, but here we only generate light rays for each individual data set. The code to do this is shown below, where all light rays are chosen to be parallel with the z-axis. Make sure that mpi4py is installed correctly before running the following scripts. Note that in these codes we use yt-3.4.

```python
# 01_make_light_ray_bw_z.py
# Purpose: to make light rays parallel with z axis
# To run it on Blue Waters:
# aprun -cc 0,2,4,6,8,10,12,14,16,18,20,22,24,26,28,30 -n 1024
# python 01_make_light_ray_bw_z.py <simulation> <file_id> --parallel

import sys
import yt
yt.enable_parallelism()
from yt.analysis_modules.cosmological_observation.light_ray.api import \
```

```python
    LightRay
import numpy as na
import mpi4py.MPI as MPI


comm = MPI.COMM_WORLD
comm_rank = comm.Get_rank()
comm_size = comm.Get_size()


sim = sys.argv[1]
file_id = na.int(sys.argv[2])
# Double check the path and make sure relevant folders are created
data_path = '/scratch/sciteam/madcpf/%s/'%sim
result_path = '/scratch/sciteam/madcpf/%s/LR/'%sim


res = 1024 # grids of the simulation box
dx  = 1.0/res # cell width in simulation units
N_sample = res*res


lr = LightRay(data_path+"Dumps/RD%04d/RD%04d"%(file_id, file_id))
num_procs = -1
count = 0 # keep track of the number of files been generated


# All light rays generated here are from axis z=0 plane to z=1 plane
# and are parallel with z axis.
# Save both H I and He II density fields for later use.
# Velecity fields are also saved when use_peculiar_velocity=True.
for i in yt.parallel_objects(xrange(N_sample), num_procs):
    i_x = na.int(i/res)
            i_y = i % res
```

```
        lr.make_light_ray(start_position = \
            [(i_x+0.5)*dx, (i_y+0.5)*dx, 0.],
                            end_position = \
            [(i_x+0.5)*dx, (i_y+0.5)*dx, 1.],
                            data_filename = result_path + \
            'lightray_RD%04d_%07d.h5'%(file_id,i),
                            fields = ['temperature',
              'H_p0_number_density',
              'He_p1_number_density'],
                            use_peculiar_velocity = True)
   count += 1


counts = comm.reduce(count, root=0, op=MPI.SUM)


# Write info into the log file after all light rays are made.
if comm_rank==0:
    f = open('/scratch/sciteam/madcpf/%s/log_analy.txt'%sim,
        'a')
    a = ''
    f.write('%45s %20s %2s %3s %8d\n'%(sys.argv[0].split('/')[-1],
            sim, a, sys.argv[2],
            na.int(counts)))
    f.close()
```

Step 2: synthetic spectra - After making light rays, we are ready to generate H I / He II Ly α spectra from them. To make sure that yt includes thermal effects, peculiar velocity and cosmic expansion in spectrum generation, we compared yt's results with Zhang's code, and found that the spectra for both H I and He II were virtually identical. Zhang's code was originally written in Fortran. We re-wrote it in Python

and it's saved as *chen-thesis/spec_analy/spec4.py*. We still use yt here since it's faster.
The code is shown below.

```python
# 02_absorption_spectrum_bw.py
# Purpose: to generate H I or He II Ly-alpha absorption spectrum
# Use after 01_make_light_ray_bw_z.py is executed
# To run it on Blue Waters:
# aprun -cc 0,2,4,6,8,10,12,14,16,18,20,22,24,26,28,30 -n 1024
# python 02_absorption_spectrum_bw.py <simulation> <ion>
# <file_id> --parallel


import yt
yt.enable_parallelism()
from yt.analysis_modules.absorption_spectrum.api import \
    AbsorptionSpectrum
import numpy as na
import sys
import mpi4py.MPI as MPI


comm = MPI.COMM_WORLD
comm_rank = comm.Get_rank()
comm_size = comm.Get_size()


# Double check the path and make sure relevant folders are created
data_path = '/scratch/sciteam/madcpf/%s/LR/'%sys.argv[1]
result_path = '/scratch/sciteam/madcpf/%s/spectrum_%s/'%(sys.argv[1],
                                                         sys.argv[2])


file_id = na.int(sys.argv[3])
# The following dicts are start and end redshifts of
```

```python
# different simulations.
# One way to get them is to check the generated light rays, e.g.
# h5ls -d lightray_RD0054_0000000.h5/grid/redshift
if sys.argv[1]=='1024_80Mpc_z4' or sys.argv[1]=='1024_80Mpc_BGonly':
    z_hs = {
        54:2.0,
        38:2.4999999861624,
        27:2.9999999204133,
        18:3.4999994722167,
        11:4.0,
        }
    z_ls = {
        54:1.92132206336084,
        38:2.40243977465138,
        27:2.88196781795282,
        18:3.36006756299821,
        11:3.83685846246863,
        }
elif sys.argv[1]=='1024_comet':
    z_hs = {
        105:2.0,
        96:2.4999998700057,
        88:2.9999999294358,
        80:3.4999998867292,
        75:3.9999996971175,
        }
    z_ls = {
        105:1.98005967875889,
        96:2.47524205747867,
```

```python
        88:2.97000927605966 ,

        80:3.46440683997484 ,

        75:3.95846202045425 ,

        }
z_h = z_hs [ file_id ]
z_l = z_ls [ file_id ]


if  sys . argv [2]== 'H ':

    my_label = 'HI  Lya '

    field = 'H_p0_number_density '

    wavelength = 1215.6700 # Angstroms

    f_value = 4.164E−01

    gamma = 6.265 e+08

    mass = 1.00794
elif  sys . argv [2]== 'He ':

    my_label = 'HeII  Lya '

    field = 'He_p1_number_density '

    wavelength = 1215.6700/4 # Angstroms

    f_value = 4.164E−01

    gamma = 100.27 e+08

    mass = 4.0026
wl_h = wavelength *(1+ z_h )
wl_l = wavelength *(1+ z_l )
res = 1024 # grids  of  the  simulation  box
# Here  we  choose  1024  pixels  for  each  spectrum ,  which  equals  with
# the  grids  of  the  simulation . An  increase  of  the  pixels  to  5000
# or  25000  does  not  significantly  alter  the  spectrum
pixels = res
sp = AbsorptionSpectrum ( wl_l ,  wl_h ,  pixels )
```

```
sp.add_line(my_label, field, wavelength, f_value,
            gamma, mass, label_threshold=1.e10)


num_procs = -1
res = 1024 # grids of the simulation box
N_sample = res*res
count = 0 # keep track of the number of files been generated


# Use use_peculiar_velocity = True only if the same flag
# is used in make_light_ray()
for i in yt.parallel_objects(xrange(N_sample), num_procs):
    sp.make_spectrum(data_path+'lightray_RD%04d_%07d.h5'%(file_id,i),
                     output_file = result_path + \
                         'spectrum_RD%04d_%07d.h5'%(file_id,i),
                     use_peculiar_velocity = True)
    count += 1
counts = comm.reduce(count, root=0, op=MPI.SUM)


# Write info into the log file after all spectra are made.
if comm_rank==0:
    f = open('/scratch/sciteam/madcpf/%s/log_analy.txt'%sys.argv[1],
             'a')
    f.write('%45s %20s %2s %3s %8d\n'%(sys.argv[0].split('/')[-1],
                                       sys.argv[1], sys.argv[2],
                                       sys.argv[3], na.int(counts)))
    f.close()
```

Step 3: non-parametric statistics - After generating spectra, we use MPI to loop over all spectra and calculate it's mean flux, flux PDF and τ PDF. We use parallelism of

MPI instead of yt since MPI is faster. The code is shown below.

```python
# 06_tau_eff_pdfs_bw.py
# Purpose: to calculate mean flux (or equivalently, tau_eff),
# flux PDF and tau PDF
# Use after 01_make_light_ray_bw_z.py and
# 02_absorption_spectrum_bw.py are executed.
# To run it on Blue Waters:
# aprun -cc 0,2,4,6,8,10,12,14,16,18,20,22,24,26,28,30 -n 1024
# python 06_tau_eff_pdfs_bw.py simulation> <ion> <file_id>
# --parallel


from mpi4py import MPI
import h5py
import numpy as na
import sys


# Here we use mpi instead of parallel processing with yt,
# to get a better speed.
comm = MPI.COMM_WORLD
comm_rank = comm.Get_rank()
comm_size = comm.Get_size()


# Double check paths and make sure relevant folders are created
data_path = '/scratch/sciteam/madcpf/%s/spectrum_%s/'%(sys.argv[1],
                                                        sys.argv[2])
result_path = '/scratch/sciteam/madcpf/%s/results/'%sys.argv[1]


file_id = na.int(sys.argv[3])
res = 1024 # grids of the simulation box
```

```python
N_sample = res*res
pixels = res # number of pixels per spectrum
# The folowing array contains the ids that each task will
# go to take care respectively
local_ids = xrange(comm_rank, N_sample, comm_size)


f_sum = 0.0
f_count = 0
bins = na.append(0,na.linspace(0.025,0.975,20))
bins = na.append(bins,1.0) # bins for flux PDF statistics
f_pdf = na.zeros(21)
bins_tau = na.linspace(-5, 3, 101) # bins for tau PDF statistics
tau_pdf = na.zeros(100)


for id in local_ids:
  try:
     file = 'spectrum_RD%04d_%07d.h5'%(file_id ,id)
     f = h5py.File(data_path+file)
     flux = f['flux'][:]
     f.close()
     # Total Flux
     f_sum += na.sum(flux)
     f_count += 1
     # Flux PDF
     p_f, x = na.histogram(flux, bins)
     f_pdf += p_f
     # Tau PDF
     tau = f['tau'][:][na.where(f['tau'][:]>0)]
     p_tau, x = na.histogram(na.log10(tau), bins_tau)
```

```python
      tau_pdf += p_tau
  # A typical reason when seeing the following error is that the
  # spectra were not generated properly
   except:
      print "#"*10, "error in 06_tau when i=%d "%id, \
          sys.argv[2], sys.argv[3]


# Sum up the flux and the counts of spectra (should be res*res)
f_sum_all = comm.reduce(f_sum, root = 0, op = MPI.SUM)
f_count_all = comm.reduce(f_count, root = 0, op = MPI.SUM)
# Sum up the Flux PDF and Tau PDF
f_pdf_all = comm.reduce(f_pdf, root=0, op=MPI.SUM)
tau_pdf_all = comm.reduce(tau_pdf, root=0, op=MPI.SUM)


# Do the following only on the root core
if comm_rank == 0:
  # To calculate the mean flux over res*res spectra
  f_mean = na.float(f_sum_all)/na.float(f_count_all)/pixels
  # To write the mean flux into a file that later will be read in by
  # 07_flux_power_bw.py for flux power spectrum calculation
  f = open('/scratch/sciteam/madcpf/%s/f_mean_%s_%s_%s.txt'%
           (sys.argv[1], sys.argv[1], sys.argv[2], sys.argv[3]),'w')
  f.write('%23.16e'%f_mean)
  f.close()
  # To normalize and write the flux PDF into a file
  # The resulted flux_pdf should sum up to 20.0
  p_f = f_pdf_all/1.0/f_pdf_all.sum()
  p_f /= 0.05
  na.savetxt(result_path+'flux_pdf_RD%04d_original_%s_new.txt' \
```

```
                  %(file_id , sys.argv[1]) , [na.linspace(0,1,21), p_f])
# To normalize and write the tau histogram into a file.
# To make a PDF plot of tau, one need to normalize p_tau
# with the bin width in LINEAR scale
p_tau = tau_pdf_all/1.0/tau_pdf_all.sum()
na.savetxt(result_path+'tau_pdf_RD%04d_original_%s.txt' \
                  %(file_id , sys.argv[1]) ,
              [(bins_tau[1:]+bins_tau[:-1])/2.0, p_tau])
# To write info into the log file
f = open('/scratch/sciteam/madcpf/%s/log_analy.txt'%sys.argv[1] ,
            'a')
f.write('%45s %20s %2s %3s %8d %23.16e\n'%
          (sys.argv[0].split('/')[-1], sys.argv[1], \
              sys.argv[2], sys.argv[3], na.int(f_count_all), \
              f_mean))
f.close()
```

Then we move on to calculate the flux power spectrum. We again use MPI to loop over all spectra. Note that we have to calculate flux power spectrum after we get the mean flux. The code we use is shown below.

```
# 07_flux_power_bw.py
# Purpose: to calculate the flux power spectrum
# Use after 01_make_light_ray_bw_z.py, 02_absorption_spectrum_bw.py,
# and 06_tau_eff_pdfs_bw.py are executed.
# To run it on Blue Waters:
# aprun -cc 0,2,4,6,8,10,12,14,16,18,20,22,24,26,28,30 -n 1024
# python 07_flux_power_bw.py <simulation> <ion> <file_id> --parallel
# See Tytler 2009 for more details about the calculation of flux
# power spectrum.
```

```python
import mpi4py.MPI as MPI
import numpy as na
import h5py
from yt.mods import *
import sys
sys.path.append('/u/sciteam/madcpf/yt/yt/utilities/')
import cosmology


# Here we use mpi instead of parallel processing with yt,
# to get a better speed.
comm = MPI.COMM_WORLD
comm_rank = comm.Get_rank()
comm_size = comm.Get_size()


# Double check paths and make sure relevant folders are created
data_path = '/scratch/sciteam/madcpf/%s/spectrum_%s/'%(sys.argv[1],
                                                        sys.argv[2])
result_path = "/scratch/sciteam/madcpf/%s/results/"%sys.argv[1]


file_id = na.int(sys.argv[3])
res = 1024 # grids of the simulation box
N_sample = res*res


# The following dicts relate the file id with its redshift.
# L_com is in the unit of comoving Mpc
if sys.argv[1]=='1024_80Mpc_z4':
    zs = {54:2.0, 38:2.5, 27:3.0, 18:3.5, 11:4.0}
    L_com = 80.0/0.677
```

```python
elif sys.argv[1]=='1024_80Mpc_BGonly':
    zs = {54:2.0, 38:2.5, 27:3.0, 18:3.5, 11:4.0}
    L_com = 80.0/0.677
elif sys.argv[1]=='1024_comet':
    zs = {105:2.0, 96:2.5, 88:3.0, 80:3.5, 75:4.0}
    L_com = 20.0/0.677
z = zs[file_id]
# Load the file generated by 06_tau_eff_bw.py,
# which contains one float
f_mean = na.loadtxt('/scratch/sciteam/madcpf/%s/f_mean_%s_%s_%s.txt'%
                    (sys.argv[1], sys.argv[1],
                     sys.argv[2], sys.argv[3]))
f_mean = na.float(f_mean)


# The following parameters are the same for the above
# three simulations
cos = cosmology.Cosmology(hubble_constant = 0.677,
                          omega_matter = 0.309,
                          omega_lambda = 0.691,
                          omega_curvature = 0.0)


H_z = cos.expansion_factor(z)*67.7   # km/s/Mpc
L_v = L_com * H_z /(1+z) # km/s
dv = L_v/1024.0 # km/s
k_v = 2*na.pi / L_v # s/km
wavenumber = k_v * na.array(range(1,int(1024/2))) # s/km


power_sum = 0
count = 0
```

```python
# The folowing array contains the ids that each task will go to
# take care respectively
local_ids = xrange(comm_rank, N_sample, comm_size)


for id in local_ids:
  try:
      file = 'spectrum_RD%04d_%07d.h5'%(file_id ,id)
      f = h5py.File(data_path+file)
      flux = f['flux'][:]
      f.close()
      flux = flux/f_mean-1
      power = na.fft.fft(flux)/L_v*dv
      power = na.abs(power)**2*L_v
      power_sum += na.log10(power)
      count += 1
  # A typical reason when seeing the following error is that the
  # spectra were not generated properly
  except:
      print "#"*10, "error in 07_flux_power when i=%d "%id, \
          sys.argv[2], sys.argv[3]


# Sum up power spectum (an array) and counts of
# files (a number) over tasks
power_sums = comm.reduce(power_sum, root=0, op=MPI.SUM)
counts = comm.reduce(count, root=0, op=MPI.SUM)


# Do the following only on the root core
if comm_rank==0:
    # To save the flux power spectrum
```

```
    power_sums /= counts

    na.savetxt(result_path+"power_RD%04d_%s.txt"%

                (file_id, sys.argv[1]),

                (wavenumber,power_sums[1:1024/2]))

    # To write info into the log file

    f = open('/scratch/sciteam/madcpf/%s/log_analy.txt'%sys.argv[1],

            'a')

    f.write('%45s %20s %2s %3s %8d\n'%(sys.argv[0].split('/')[-1],

                                        sys.argv[1], sys.argv[2],

                                        sys.argv[3], na.int(counts)))

    f.close()
```

Step 4: parametric statistics - Another thing we could do with artificial spectra is to fit them with Voigt profile and do Doppler parameter b / column density N statistics. We use the following three scripts to do this. Since the current yt (yt-3.4) algorithm for spectra fitting gives unsatisfactory results (much more b with values equal to b_{init}), we choose to use autovp to fit H I and He II Lyα lines. This is much more time consuming but gives better results. The following script creates a folder for each task and copy autovp files into it. It also reads in spectra generated by yt and convert them to the format that autovp requires. Note that we need to add He II related parameters into *autovp/ions.dat* before running the following script.

```
# 03_prep_autovp_bw.py
# Purpose: to prepare for the fit of spectra
# Use after 01_make_light_ray_bw_z.py and 02_absorption_spectrum_bw.py
# To run it on Blue Waters:
# aprun -cc 0,2,4,6,8,10,12,14,16,18,20,22,24,26,28,30 -n 1024
# python 03_prep_autovp_bw.py <simulation> <ion> <file_id> --parallel
```

```python
from mpi4py import MPI
import h5py
import numpy as na
import sys
import os


comm = MPI.COMM_WORLD
comm_rank = comm.Get_rank()
comm_size = comm.Get_size()


# Double check paths and make sure relevant folders are created
data_path = '/scratch/sciteam/madcpf/%s/spectrum_%s/'%(sys.argv[1],
                                                       sys.argv[2])
result_path = '/scratch/sciteam/madcpf/' + \
    '%s/autovp_fitted_%s/'%(sys.argv[1], sys.argv[2])


file_id = na.int(sys.argv[3])
if sys.argv[2]=='H':
  wavelength_0 = 1215.6700 # Angstroms
elif sys.argv[2]=='He':
  wavelength_0 = 1215.6700/4 # Angstroms
speed_of_light = 2.99792458e5  # km/s


# For each task, create a folder and copy autovp files into it.
if not os.path.isdir(result_path+"%04d"%comm_rank):
  os.mkdir(result_path+"%04d"%comm_rank)
  command = 'cp /u/sciteam/madcpf/autovp/move/* %s'%result_path + \
      "%04d"%comm_rank
  os.system(command)
```

```python
res = 1024 # grids of the simulation box
N_sample = res*res
# The folowing array contains the ids that each task will go to
# take care respectively
local_ids = xrange(comm_rank, N_sample, comm_size)
count = 0 # keep track of the number of files been generated


# Read in spectra and convert it to the format that autovp requires.
for i in local_ids:
    f = h5py.File(data_path+'spectrum_RD%04d_%07d.h5'%(file_id,i))
    flux = f['flux'][:]
    wavelength = f["wavelength"][:]
    f.close()
    z0 = wavelength[0]/wavelength_0-1
    velocity = (wavelength/wavelength_0/(1+z0)-1) * speed_of_light
    noise = na.random.uniform(0.025, 0.040, len(flux))
    f_out = open(result_path+'%04d/autovp_fit_RD%04d_%07d.cln' \
                  %(comm_rank,file_id,i), 'w')
    for l in xrange(1,len(flux)):
        f_out.write("%.4f   %.4f   %.6f   %.6f\n"%(wavelength[l],
                                          velocity[l], flux[l], noise[l]))
    f_out.close()
    count += 1
counts = comm.reduce(count, root=0, op=MPI.SUM)


# Write info into the log file.
if comm_rank==0:
    f = open('/scratch/sciteam/madcpf/%s/log_analy.txt' \
```

```
                %sys . argv [ 1 ] , ' a ' )
    f . write ( '%45s  %20s  %2s  %3s  %8d\n'%(sys . argv [ 0 ] . split ( '/' ) [ −1 ] ,
                sys . argv [ 1 ] ,  sys . argv [ 2 ] ,  sys . argv [ 3 ] ,  na . int ( counts ) ) )
    f . close ( )
```

Then we use the following script to fit spectra with autovp. The input file and
output file have fixed name defined by autovp, so we need to rename files before and
after each fit.

```
# 04 _fitting_spectrum_autovp_bw . py
# Purpose :  to  fit  spectra  using  autovp
# Use  after  01 _make_light_ray_bw_z . py ,  02 _absorption_spectrum_bw . py ,
# and  03 _prep_autovp_bw . py
# To  run  it  on  Blue  Waters :
# aprun  −cc  0 ,2 ,4 ,6 ,8 ,10 ,12 ,14 ,16 ,18 ,20 ,22 ,24 ,26 ,28 ,30  −n  1024
# python  04 _fitting_spectrum_autovp_bw . py  <simulation> <ion>
# <file_id > −−parallel


from  mpi4py  import  MPI
import  sys
import  numpy  as  na
import  os


comm  =  MPI . COMM_WORLD
comm_rank  =  comm . Get_rank ( )
comm_size  =  comm . Get_size ( )


# Double  check  paths  and  make  sure  relevant  folders  are  created
data_path  =  '/scratch/sciteam/madcpf/%s/autovp_fitted_%s/ '  \
                                        %( sys . argv [ 1 ] , sys . argv [ 2 ] )
```

```python
file_id = na.int(sys.argv[3])
res = 1024 # grids of the simulation box
N_sample = res*res


# For the species we are going to fit, define the commands
# to fit spectra using autovp and rename the resulted file
if sys.argv[2]=='H':
  cm = "cd %s/%04d; /bin/cp %s.cln H1216N.cln; ./autofit H1216N;" +\
       " ./minfit H1216N.pro; /bin/mv H1216N.res %s.res;" +\
       " /bin/mv H1216N.vpm %s.vpm"
elif sys.argv[2]=='He':
  cm = "cd %s/%04d; /bin/cp %s.cln HeII304.cln; ./autofit HeII304;" +\
       " ./minfit HeII304.pro; /bin/mv HeII304.res %s.res;" +\
       " /bin/mv HeII304.vpm %s.vpm"


# The folowing array contains the ids that each task will go to
# take care respectively
local_ids = xrange(comm_rank, N_sample, comm_size)
count = 0 # keep track of the number of files been generated


# Loop over all spetra to execute the command defined above
for i in local_ids:
  file = 'autovp_fit_RD%04d_%07d'%(file_id, i)
  command = cm%(data_path, comm_rank, file, file, file)
  os.system(command)
  count += 1


counts = comm.reduce(count, root=0, op=MPI.SUM)
```

```
# Write info into the log file.
if comm_rank==0:
    f = open('/scratch/sciteam/madcpf/%s/log_analy.txt'%sys.argv[1],'a')
    f.write('%45s %20s %2s %3s %8d\n'%(sys.argv[0].split('/')[-1],
            sys.argv[1], sys.argv[2], sys.argv[3], na.int(counts)))
    f.close()
```

Finally we could loop over all the fitting results, read N and b of each absorber and do statistics of them. In the following script we calculate the histogram of b and fit it with Hui-Rutledge function (Hui and Rutledge, 1999).

```
# 05_analyze_LAF_autovp_bw.py
# Purpose: to do N / b statistics of the fitting results
# Use after 01_make_light_ray_bw_z.py, 02_absorption_spectrum_bw.py,
# 03_prep_autovp_bw.py and 04_fitting_spectrum_autovp_bw.py
# To run it on Blue Waters:
# aprun -cc 0,2,4,6,8,10,12,14,16,18,20,22,24,26,28,30 -n 1024
# python 05_analyze_LAF_autovp_bw.py <simulation> <ion> <file_id>
# --parallel


import matplotlib
matplotlib.use('Agg')
import matplotlib.pyplot as plt
from mpi4py import MPI
import sys
import numpy as na


# Double check paths and make sure relevant folders are created
data_path = '/scratch/sciteam/madcpf/%s/autovp_fitted_%s/' \
```

```python
                                            %(sys.argv[1], sys.argv[2])
result_path = "/scratch/sciteam/madcpf/%s/results/"%sys.argv[1]


comm = MPI.COMM_WORLD
comm_rank = comm.Get_rank()
comm_size = comm.Get_size()


file_id = na.int(sys.argv[3])


# Find the redshift corresponding to the current data dump
if sys.argv[1]=='1024_80Mpc_z4' or sys.argv[1]=='1024_80Mpc_BGonly':
    rss = {54:2.0, 38:2.5, 27:3.0, 18:3.5, 11:4.0}
elif sys.argv[1]=='1024_comet':
    rss = {105:2.0, 96:2.5, 88:3.0, 80:3.5, 75:4.0}
rs = rss[file_id]


res = 1024 # grids of the simulation box
N_sample = res*res


input_fields = ['N', 'b']
field_data = {}
for field in input_fields:
    field_data[field] = []


# The folowing array contains the ids that each task will go to
# take care respectively
local_ids = xrange(comm_rank, 1024*1024, comm_size)
count = 0 # keep track of the number of files been generated
for i in local_ids:
```

```python
    file = 'autovp_fit_RD%04d_%07d.vpm'%(file_id,i)
    N, b = na.loadtxt(data_path+"%04d/"%comm_rank+file, skiprows=2,
                        usecols=(1,3), unpack=True)
    # Here we specify a range of N that we are going to analyze
    ind = na.logical_and(N>na.power(10, -0.5), N<na.power(10, 1.5))
    field_data['N'] = na.append(field_data['N'], N[ind])
    field_data['b'] = na.append(field_data['b'], b[ind])
    count += 1


# To gather all N's and b's and to sum up the count of files
Ns = comm.gather(field_data['N'].tolist(),root=0)
bs = comm.gather(field_data['b'].tolist(),root=0)
counts = comm.reduce(count, root=0, op=MPI.SUM)
del field_data


# Define the Hui-Rutledge function to fit distribution of b
from scipy.optimize import curve_fit
def f_HR(x, B, b_sigma):
    return B*b_sigma**4/x**5*na.exp(-b_sigma**4/x**4)


if comm_rank == 0:
    # Here we specify a range of b that we are going to analyze
    ind = na.logical_and(bs>0, bs<999)
    # Get the PDF of b
    counts, edges = na.histogram(bs[ind], na.array(range(0,80,2)))
    # Fit b with Hui-Rutlidge function
    popt6, pcov6 = curve_fit(f_HR, range(1,79,2),
                                counts*1.0/counts.sum()/2)
    # Save the b PDF and fitting results into a file
```

```python
xdata = na.linspace(0.1,78,200)
na.save(result_path+'%s_f_b_vs_b_%.1f_autovp_cut_N_fit' \
            %(sys.argv[1], rs),
        [counts, edges, xdata, f_HR(xdata,popt6[0],popt6[1])])
# Write info into the log file.
f = open('/scratch/sciteam/madcpf/%s/log_analy.txt'%sys.argv[1],'a')
f.write('%45s %20s %2s %3s %8d\n'%(sys.argv[0].split('/')[-1],
        sys.argv[1], sys.argv[2], sys.argv[3], na.int(counts)))
f.close()
# Make plots of the distribution with HR function.
# This could also be done locally.
i_max = na.where(counts==counts.max())[0][0]
plt.hist(bs[ind], edges, histtype='bar', normed=1)
plt.plot(xdata, f_HR(xdata,popt6[0],popt6[1]),'r')
plt.xlabel('$b [km/s]$')
plt.ylabel('$f(b)$')
plt.title(r'%s, $z=%.1f$, $b_{max}=%d km/s$, ' + \
            '$b_\sigma=%.1f km/s$'%(sys.argv[1],
                    rs,round(edges[i_max]+1),popt6[1]))
plt.savefig(result_path+'%s_f_b_vs_b_%.1f_autovp_cut_N.png' \
            %(sys.argv[1], rs))
plt.clf()
```

A.3.2   Comet

To do spectral analysis to all light rays in z direction, we need to read and write millions of files in our case, which would make lustre file system unstable. However, the latency to SSDs is much shorter and thus make them ideal for fast I/O. So we

rewrite the codes for spectral analysis to use SSD scratch space in Comet. Due to the fact that the files generated in SSDs are erased after the job is done, we write files in SSDs (or copy from existed tar files from lustre and extract them in SSDs), do analysis there, tar the output files, and move them to lustre. Since different SSD nodes don't share memory, we also need to make sure that after we create (or extract) files in SSDs in different nodes, cores in the same node will analyze them. The scripts are saved under *chen-thesis/spec_analy/comet*, and are briefly discussed below.

c1_lr_spec.py creates light rays and absorption spectra. Its function combines *01_make_light_ray_bw_z.py* and *02_absorption_spectrum_bw.py* in Blue Waters.

c2_spec_analy.py calculates mean flux, flux PDF and τ PDF, and it's like *06_tau_eff_pdfs_bw.py*.

c2_spec_analy_sample.py draws 10 samples, each with $\sim 10^4$ spectra, and calculates their mean flux, flux PDF and τ PDF. It's purpose is to show the variation of those quantities across the simulation box.

c3_power.py calculates flux power spectrum, and it's like *07_flux_power_bw.py*.

c4_fit.py fits spectra with autovp, and it's like the combination of *03_prep_autovp-_bw.py*, *04_fitting_spectrum_autovp_bw.py*, and *05_analyze_LAF_autovp_bw.py*.

Bibliography

Abel, T., Anninos, P., Zhang, Y., and Norman, M. L. (1997). Modeling primordial gas in numerical cosmology. *New A.*, 2:181–207.

Abel, T. and Haehnelt, M. G. (1999). Radiative Transfer Effects during Photoheating of the Intergalactic Medium. *The Astrophysical Journal Letters*, 520:L13–L16.

Abel, T. and Hähnelt, M. G. (1999). Radiative Transfer Effects during Photoheating of the Intergalactic Medium. *Astrophys.J.*, 520(astro-ph/9903102. 1, pt.2):L13.

Ahn, K., Iliev, I. T., Shapiro, P. R., Mellema, G., Koda, J., and Mao, Y. (2012). Detecting the Rise and Fall of the First Stars by Their Impact on Cosmic Reionization. *The Astrophysical Journal Letters*, 756:L16.

Anninos, P., Zhang, Y., Abel, T., and Norman, M. L. (1997). Cosmological hydrodynamics with multi-species chemistry and nonequilibrium ionization and cooling. *New Astronomy*, 2:209–224.

Becker, G. D., Bolton, J. S., Haehnelt, M. G., and Sargent, W. L. W. (2011). Detection of extended He II reionization in the temperature evolution of the intergalactic medium. *Monthly Notices of the Royal Astronomical Society*, 410(2):1096–1112.

Becker, G. D., Bolton, J. S., Madau, P., Pettini, M., Ryan-Weber, E. V., and Venemans, B. P. (2014). Evidence of patchy hydrogen reionization from an extreme Lyα trough below redshift six. *arXiv.org*, page arXiv:1407.4850.

Becker, G. D., Hewett, P. C., Worseck, G., and Prochaska, J. X. (2012). A Refined Measurement of the Mean Transmitted Flux in the Ly-alpha Forest over 2 < z < 5 Using Composite Quasar Spectra. *arXiv.org*, page arXiv:1208.2584.

Becker, G. D., Rauch, M., and Sargent, W. L. W. (2007). The Evolution of Optical Depth in the Lyα Forest: Evidence Against Reionization at z~6. *The Astrophysical Journal*, 662(1):72–93.

Becker, R. H., Fan, X., White, R. L., Strauss, M. A., Narayanan, V. K., Lupton, R. H., Gunn, J. E., Annis, J., Bahcall, N. A., Brinkmann, J., Connolly, A. J., Csabai, I., Czarapata, P. C., Doi, M., Heckman, T. M., Hennessy, G. S., Ivezic, Z., Knapp, G. R., Lamb, D. Q., McKay, T. A., Munn, J. A., Nash, T., Nichol, R.,

Pier, J. R., Richards, G. T., Schneider, D. P., Stoughton, C., Szalay, A. S., Thakar, A. R., and York, D. G. (2001). Evidence for Reionization at z ˜ 6: Detection of a Gunn-Peterson Trough in a z=6.28 Quasar. *arXiv.org*.

Boera, E., Murphy, M. T., Becker, G. D., and Bolton, J. S. (2014). The thermal history of the intergalactic medium down to redshift z = 1.5: a new curvature measurement. *ArXiv e-prints*, 441(3):1916–1933.

Bolton, J. S., Becker, G. D., Haehnelt, M. G., and Viel, M. (2013). A consistent determination of the temperature of the intergalactic medium at redshift z=2.4. *arXiv.org*.

Bolton, J. S., Becker, G. D., Raskutti, S., Wyithe, J. S. B., Haehnelt, M. G., and Sargent, W. L. W. (2012). Improved measurements of the intergalactic medium temperature around quasars: possible evidence for the initial stages of He II reionization at z 6. *Monthly Notices of the Royal Astronomical Society*, 419(4):2880–2892.

Bolton, J. S., Oh, S. P., and Furlanetto, S. R. (2009). The evolution of the Lyα forest effective optical depth following HeII reionization. *Monthly Notices of the Royal Astronomical Society*, 396(4):2405–2418.

Bryan, G. L., Machacek, M., Anninos, P., and Norman, M. L. (1999). Resolving the LYalpha Forest. *The Astrophysical Journal*, 517:13–30.

Bryan, G. L., Norman, M. L., O'Shea, B. W., Abel, T., Wise, J. H., Turk, M. J., Reynolds, D. R., Collins, D. C., Wang, P., Skillman, S. W., Smith, B., Harkness, R. P., Bordner, J., Kim, J.-h., Kuhlen, M., Xu, H., Goldbaum, N., Hummels, C., Kritsuk, A. G., Tasker, E., Skory, S., Simpson, C. M., Hahn, O., Oishi, J. S., So, G. C., Zhao, F., Cen, R., Li, Y., and The Enzo Collaboration (2014a). ENZO: An Adaptive Mesh Refinement Code for Astrophysics. *The Astrophysical Journal Supplement Series*, 211:19.

Bryan, G. L., Norman, M. L., O'Shea, B. W., Abel, T., Wise, J. H., Turk, M. J., Reynolds, D. R., Collins, D. C., Wang, P., Skillman, S. W., Smith, B., Harkness, R. P., Bordner, J., Kim, J.-h., Kuhlen, M., Xu, H., Goldbaum, N., Hummels, C., Kritsuk, A. G., Tasker, E., Skory, S., Simpson, C. M., Hahn, O., Oishi, J. S., So, G. C., Zhao, F., Cen, R., Li, Y., and The Enzo Collaboration (2014b). ENZO: An Adaptive Mesh Refinement Code for Astrophysics. *The Astrophysical Journal Supplement Series*, 211:19.

Bryan, G. L., Norman, M. L., Stone, J. M., Cen, R., and Ostriker, J. P. (1995). A piecewise parabolic method for cosmological hydrodynamics. *Computer Physics Communications*, 89:149–168.

Carswell, R. F., Morton, D. C., Smith, M. G., Stockton, A. N., Turnshek, D. A., and

Weymann, R. J. (1984). The absorption line profiles in Q1101-264. *Astrophysical Journal*, 278:486–498.

Cen, R., Miralda-Escudé, J., Ostriker, J. P., and Rauch, M. (1994). Gravitational Collapse of Small-Scale Structure as the Origin of the Lyman Alpha Forest. *arXiv.org*, pages arXiv:math–9501232.

Choudhury, T. R., Ferrara, A., and Gallerani, S. (2008). On the minimum mass of reionization sources. *Monthly Notices of the Royal Astronomical Society*, 385:L58–L62.

Colella, P. and Woodward, P. R. (1984). The Piecewise Parabolic Method (PPM) for gas-dynamical simulations. *J. Comp. Phys.*, 54:174–201.

Compostella, M., Cantalupo, S., and Porciani, C. (2013). The imprint of inhomogeneous He II reionization on the H I and He II Lyα forest. *Monthly Notices of the Royal Astronomical Society*, 435(4):3169–3190.

Dall'Aglio, A., Wisotzki, L., and Worseck, G. (2009). The UV background photoionization rate at $2.3 < z < 4.6$ as measured from the Sloan Digital Sky Survey. *arXiv.org*, page arXiv:0906.1484.

Davé, R., Hernquist, L., Weinberg, D. H., and Katz, N. (1997). Voigt-Profile Analysis of the Lyα Forest in a Cold Dark Matter Universe. *The Astrophysical Journal*, 477(1):21–26.

Day, A. (2013). *An Analysis of Astrophysics and Fundamental Physics from the Lyman-alpha Forest*. PhD thesis, University of California, San Diego.

Dixon, K. L. and Furlanetto, S. R. (2009). The evolution of the helium-ionizing background at z ˜ 2-3. *ArXiv e-prints*, 706(2):970–979.

Fan, X., Strauss, M. A., Becker, R. H., White, R. L., Gunn, J. E., Knapp, G. R., Richards, G. T., Schneider, D. P., Brinkmann, J., and Fukugita, M. (2005). Constraining the Evolution of the Ionizing Background and the Epoch of Reionization with z ˜ 6 Quasars II: A Sample of 19 Quasars. *arXiv.org*.

Faucher-Giguere, C. A., Prochaska, J. X., Lidz, A., Hernquist, L., and Zaldarriaga, M. (2008). A Direct Precision Measurement of the Intergalactic Lyman-alpha Opacity at 2. *ArXiv e-prints*, 681(2):831–855.

Fechner, C., Reimers, D., Kriss, G. A., Baade, R., Blair, W. P., Giroux, M. L., Green, R. F., Moos, H. W., Morton, D. C., Scott, J. E., Shull, J. M., Simcoe, R., Songaila, A., and Zheng, W. (2006). The UV spectrum of HS 1700+6416. II. FUSE observations of the HeII Lyman alpha forest. *Astronomy and Astrophysics*, 455(1):91–106.

Giallongo, E., Grazian, A., Fiore, F., Fontana, A., Pentericci, L., Vanzella, E., Dickinson, M., Kocevski, D., Castellano, M., Cristiani, S., Ferguson, H., Finkelstein, S., Grogin, N., Hathi, N., Koekemoer, A. M., Newman, J. A., and Salvato, M. (2015). Faint AGNs at z > 4 in the CANDELS GOODS-S field: looking for contributors to the reionization of the Universe. *Astronomy and Astrophysics*, 578:A83.

Greenstein, J. L. and Schmidt, M. (1964). The Quasi-Stellar Radio Sources 3c 48 and 3c 273. *Astrophysical Journal*, 140:1.

Gustafson, S. (2016). *Measurements of Primordial Deuterium and Lyman-Alpha Forest Evolution with High-Resolution Quasar Spectroscopy*. PhD thesis, University of California, San Diego.

Haardt, F. and Madau, P. (1995). Radiative Transfer in a Clumpy Universe: II. The Utraviolet Extragalactic Background. *arXiv.org*, page 20.

Haardt, F. and Madau, P. (2012). Radiative transfer in a clumpy universe: IV. New synthesis models of the cosmic UV/X-ray background. *Astrophys.J.*, 746:125.

Hayes, J. C. and Norman, M. L. (2003). Beyond Flux-limited Diffusion: Parallel Algorithms for Multidimensional Radiation Hydrodynamics. *Ap. J. Supp.*, 147:197–220.

Hockney, R. W. and Eastwood, J. W. (1988). *Computer simulation using particles*. CRC Press.

Hopkins, P. F., Richards, G. T., and Hernquist, L. (2006). An Observational Determination of the Bolometric Quasar Luminosity Function. *arXiv.org*.

Hui, L. and Gnedin, N. Y. (1997). Equation of state of the photoionized intergalactic medium. *MNRAS*, 292:27–42.

Hui, L. and Haiman, Z. (2003). The thermal memory of reionization history. *The Astrophysical Journal*, 596(1):9.

Hui, L. and Rutledge, R. E. (1999). The b distribution and the velocity structure of absorption peaks in the Lyman alpha forest. *ArXiv e-prints*, 517(2):541–548.

Iliev, I. T., Ciardi, B., Alvarez, M. A., Maselli, A., Ferrara, A., Gnedin, N. Y., Mellema, G., Nakamoto, T., Norman, M. L., Razoumov, A. O., Rijkhorst, E.-J., Ritzerveld, J., Shapiro, P. R., Susa, H., Umemura, M., and Whalen, D. J. (2006). Cosmological radiative transfer codes comparison project - I. The static density field tests. *MNRAS*, 371:1057–1086.

Iliev, I. T., Mellema, G., Ahn, K., Shapiro, P. R., Mao, Y., and Pen, U.-L. (2014). Simulating cosmic reionization: how large a volume is large enough? *Monthly Notices of the Royal Astronomical Society*, 439:725–743.

Iliev, I. T., Mellema, G., Shapiro, P. R., and Pen, U. (2007). Self-regulated reionization. *Monthly Notices of the Royal Astronomical Society*, 376:534–548.

Iliev, I. T., Mellema, G., Shapiro, P. R., Pen, U.-L., Mao, Y., Koda, J., and Ahn, K. (2012). Can 21-cm observations discriminate between high-mass and low-mass galaxies as reionization sources? *Monthly Notices of the Royal Astronomical Society*, 423:2222–2253.

Jena, T., Norman, M. L., Tytler, D., Kirkman, D., Suzuki, N., Chapman, A., Melis, C., Paschos, P., O'Shea, B., So, G., Lubin, D., Lin, W.-C., Reimers, D., Janknecht, E., and Fechner, C. (2005). A concordance model of the Lyman α forest at z= 1.95. *Monthly Notices of the Royal Astronomical Society*, 361(1):70–96.

Kauffmann, G. and Haehnelt, M. (2000). A unified model for the evolution of galaxies and quasars. *Monthly Notices of the Royal Astronomical Society*, 311:576–588.

Kim, T. S., Bolton, J. S., Viel, M., Haehnelt, M. G., and Carswell, R. F. (2007). An improved measurement of the flux distribution of the Ly-alpha forest in QSO absorption spectra: the effect of continuum fitting, metal contamination and noise properties. *arXiv.org*.

La Plante, P., Trac, H., Croft, R., and Cen, R. (2016). Helium Reionization Simulations. II. Signatures of Quasar Activity on the IGM. *arXiv.org*.

Lukić, Z., Stark, C. W., Nugent, P., White, M., Meiksin, A. A., and Almgren, A. (2015). The Lyman α forest in optically thin hydrodynamical simulations. *Monthly Notices of the Royal Astronomical Society*, 446(4):3697–3724.

Madau, P. and Meiksin, A. (1994). The He II Lyman-alpha opacity of the universe. *Astrophysical Journal*, 433:L53–L56.

Masters, D. and others (2012). Evolution of the Quasar Luminosity Function Over 3 . *ArXiv e-prints*, 755:169.

McDonald, P., Miralda-Escudé, J., Rauch, M., Sargent, W. L. W., Barlow, T. A., and Cen, R. (2001). A Measurement of the Temperature-Density Relation in the Intergalactic Medium Using a New Lyα Absorption-Line Fitting Method. *The Astrophysical Journal*, 562(1):52–75.

McDonald, P., Miralda-Escudé, J., Rauch, M., Sargent, W. L. W., Barlow, T. A., Cen, R., and Ostriker, J. P. (2000). The Observed Probability Distribution Function, Power Spectrum, and Correlation Function of the Transmitted Flux in the Lyα Forest. *The Astrophysical Journal*, 543(1):1–23.

McDonald, P., Seljak, U., Burles, S., Schlegel, D. J., Weinberg, D. H., Cen, R., Shih, D., Schaye, J., Schneider, D. P., Bahcall, N. A., Briggs, J. W., Brinkmann, J., Brunner, R. J., Fukugita, M., Gunn, J. E., Ivezić, Ž., Kent, S., Lupton, R. H., and

Vanden Berk, D. E. (2006). The Lyα Forest Power Spectrum from the Sloan Digital Sky Survey. *The Astrophysical Journal Supplement Series*, 163:80–109.

McDonald, P., Seljak, U., Cen, R., Shih, D., Weinberg, D. H., Burles, S., Schneider, D. P., Schlegel, D. J., Bahcall, N. A., Briggs, J. W., Brinkmann, J., Fukugita, M., Ivezic, Z., Kent, S., and Vanden Berk, D. E. (2005). The Linear Theory Power Spectrum from the Lyα Forest in the Sloan Digital Sky Survey. *The Astrophysical Journal*, 635(2):761–783.

McGreer, I. D. and others (2013). The z=5 Quasar Luminosity Function from SDSS Stripe 82. *ArXiv e-prints*, 768:105.

McQuinn, M., Lidz, A., Zaldarriaga, M., Hernquist, L., Hopkins, P. F., Dutta, S., and Faucher-Giguère, C.-A. (2009). He II Reionization and its Effect on the Intergalactic Medium. *The Astrophysical Journal*, 694:842–866.

Meiksin, A. and Tittley, E. R. (2012). The impact of helium reionization on the structure of the intergalactic medium. *ArXiv e-prints*, 423:7.

Miralda-Escudé, J., Haehnelt, M., and Rees, M. J. (2000). Reionization of the Inhomogeneous Universe. *The Astrophysical Journal*, 530:1–16.

Morel, J. E. (2000). Diffusion-limit asymptotics of the transport equation, the $p_{1/3}$ equations, and two flux-limited diffusion theories. *J. Quant. Spectrosc. Radiat. Transfer*, 65:769–778.

Norman, M. L. and Bryan, G. L. (1999). Cosmological adaptive mesh refinement. In Miyama, S. M., Tomisaka, K., and Hanawa, T., editors, *Numerical Astrophysics*, volume 240 of *Astrophysics and Space Science Library*, page 19.

Norman, M. L., Bryan, G. L., Harkness, R., Bordner, J., Reynolds, D. R., O'Shea, B., and Wagner, R. (2007). *Petascale Computing: Algorithms and Applications*, chapter Simulating Cosmological Evolution with *Enzo*. CRC Press.

Norman, M. L., Reynolds, D. R., So, G. C., Harkness, R. P., and Wise, J. H. (2015). Fully Coupled Simulation of Cosmic Reionization. I. Numerical Methods and Tests. *The Astrophysical Journal Supplement Series*, 216:16.

Oke, J. B. and Korycansky, D. G. (1982). Absolute spectrophotometry of very large redshift quasars. *Astrophysical Journal*, 255:11–19.

O'Meara, J. M., Tytler, D., Kirkman, D., Suzuki, N., Prochaska, J. X., Lubin, D., and Wolfe, A. M. (2001). The Deuterium to Hydrogen Abundance Ratio toward a Fourth QSO: HS 0105+1619. *The Astrophysical Journal*, 552:718–730.

O'Shea, B. W., Bryan, G., Bordner, J. O., Norman, M. L., Abel, T., Harkness, R., and Kritsuk, A. (2004). *Adaptive Mesh Refinement – Theory and Applications*,

chapter Introducing *Enzo*, an AMR Cosmology Application. Lecture Notes in Computational Science and Engineering. Springer.

O'Shea, B. W., Wise, J. H., Xu, H., and Norman, M. L. (2015). Probing the Ultraviolet Luminosity Function of the Earliest Galaxies with the Renaissance Simulations. *The Astrophysical Journal Letters*, 807:L12.

Osterbrock, D. E. (1989). *Astrophysics of gaseous nebulae and active galactic nuclei.* University Science Books.

Paschos, P. (2005). *On the Ionization and Chemical Evolution of the Intergalactic Medium.* PhD thesis, University of Illinois at Urbana-Champaign.

Paschos, P., Norman, M. L., Bordner, J. O., and Harkness, R. (2007). Late Reheating of the IGM by Quasars: A Radiation Hydrodynamical Simulation of Helium II Reionization. *ArXiv e-prints.*

Pawlik, A. H., Schaye, J., and van Scherpenzeel, E. (2009). Keeping the Universe ionized: photoheating and the clumping factor of the high-redshift intergalactic medium. *Monthly Notices of the Royal Astronomical Society*, 394(4):1812–1824.

Puchwein, E., Bolton, J. S., Haehnelt, M. G., Madau, P., Becker, G. D., and Haardt, F. (2014). The photoheating of the intergalactic medium in synthesis models of the UV background. *arXiv.org.*

Rauch, M. (1998). The Lyman Alpha Forest in the Spectra of QSOs. *arXiv.org.*

Reimers, D., Fechner, C., Hagen, H. J., Jakobsen, P., Tytler, D., and Kirkman, D. (2005). Intergalactic HeII absorption towards QSO 1157+3143. *Astronomy and Astrophysics*, 442(1):63–68.

Reynolds, D. R., Hayes, J. C., Paschos, P., and Norman, M. L. (2009). Self-consistent solution of cosmological radiation-hydrodynamics and chemical ionization. *J. Comput. Phys.*, 228:6833–6854.

Richards, G. T. and others (2006). The Sloan Digital Sky Survey Quasar Survey: Quasar Luminosity Function from Data Release 3. *ArXiv e-prints*, 131:2766–2787.

Ricotti, M., Gnedin, N. Y., and Shull, J. M. (2002). The Fate of the First Galaxies. II. Effects of Radiative Feedback. *The Astrophysical Journal*, 575:49–67.

Ross, N. P. and others (2013). The SDSS-III Baryon Oscillation Spectroscopic Survey: The Quasar Luminosity Function from Data Release Nine. *ArXiv e-prints*, 773:14.

Sargent, W. L. W., Young, P. J., Boksenberg, A., and Tytler, D. (1980). The distribution of Lyman-alpha absorption lines in the spectra of six QSOs - Evidence for an intergalactic origin. *Astrophysical Journal Supplement Series*, 42:41–81.

Schaye, J., Theuns, T., Rauch, M., Efstathiou, G., and Sargent, W. L. W. (2000). The Thermal history of the intergalactic medium. *ArXiv e-prints*, 318:817.

Shankar, F. and Mathur, S. (2007). On the Faint End of the High-Redshift Active Galactic Nucleus Luminosity Function. *The Astrophysical Journal*, 660(2):1051–1059.

Shapiro, P. R., Iliev, I. T., Mellema, G., Ahn, K., Mao, Y., Friedrich, M., Datta, K., Park, H., Komatsu, E., Fernandez, E., Koda, J., Bovill, M., and Pen, U.-L. (2012). Simulating cosmic reionization and the radiation backgrounds from the epoch of reionization. In Umemura, M. and Omukai, K., editors, *American Institute of Physics Conference Series*, volume 1480 of *American Institute of Physics Conference Series*, pages 248–260.

Shull, M., France, K., Danforth, C., Smith, B., and Tumlinson, J. (2010). Hubble/COS Observations of the Quasar HE 2347-4342: Probing the Epoch of He II Patchy Reionization at Redshifts z = 2.4-2.9. *arXiv.org*.

Skory, S., Turk, M. J., Norman, M. L., and Coil, A. L. (2010). Parallel HOP: A Scalable Halo Finder for Massive Cosmological Data Sets. *The Astrophysical Journal Supplement Series*, 191:43–57.

So, G. C., Norman, M. L., Reynolds, D. R., and Wise, J. H. (2014). Fully Coupled Simulation of Cosmic Reionization. II. Recombinations, Clumping Factors, and the Photon Budget for Reionization. *The Astrophysical Journal*, 789:149.

Syphers, D. and Shull, J. M. (2014). Hubble Space Telescope/Cosmic Origins Spectrograph Observations of the Quasar Q0302-003: Probing the He II Reionization Epoch and QSO Proximity Effects. *The Astrophysical Journal*, 784(1):42.

Telfer, R. C., Zheng, W., Kriss, G. A., and Davidsen, A. F. (2002). The Rest-frame extreme ultraviolet spectral properties of QSOS. *ArXiv e-prints*, 565(2):773–785.

the Planck Collaboration, P., Ade, P. A. R., Aghanim, N., Arnaud, M., Ashdown, M., Aumont, J., Baccigalupi, C., Banday, A. J., Barreiro, R. B., Bartlett, J. G., Bartolo, N., Battaner, E., Battye, R., Benabed, K., Benoit, A., Benoit-Levy, A., Bernard, J. P., Bersanelli, M., Bielewicz, P., Bonaldi, A., Bonavera, L., Bond, J. R., Borrill, J., Bouchet, F. R., Boulanger, F., Bucher, M., Burigana, C., Butler, R. C., Calabrese, E., Cardoso, J. F., Catalano, A., Challinor, A., Chamballu, A., Chary, R. R., Chiang, H. C., Chluba, J., Christensen, P. R., Church, S., Clements, D. L., Colombi, S., Colombo, L. P. L., Combet, C., Coulais, A., Crill, B. P., Curto, A., Cuttaia, F., Danese, L., Davies, R. D., Davis, R. J., de Bernardis, P., de Rosa, A., de Zotti, G., Delabrouille, J., Desert, F. X., Di Valentino, E., Dickinson, C., Diego, J. M., Dolag, K., Dole, H., Donzelli, S., Dore, O., Douspis, M., Ducout, A., Dunkley, J., Dupac, X., Efstathiou, G., Elsner, F., Ensslin, T. A., Eriksen, H. K., Farhang, M., Fergusson, J., Finelli, F., Forni, O., Frailis, M., Fraisse, A. A., Franceschi, E., Frejsel,

A., Galeotta, S., Galli, S., Ganga, K., Gauthier, C., Gerbino, M., Ghosh, T., Giard, M., Giraud-Heraud, Y., Giusarma, E., Gjerlow, E., Gonzalez-Nuevo, J., Górski, K. M., Gratton, S., Gregorio, A., Gruppuso, A., Gudmundsson, J. E., Hamann, J., Hansen, F. K., Hanson, D., Harrison, D. L., Helou, G., Henrot-Versille, S., Hernandez-Monteagudo, C., Herranz, D., Hildebrandt, S. R., Hivon, E., Hobson, M., Holmes, W. A., Hornstrup, A., Hovest, W., Huang, Z., Huffenberger, K. M., Hurier, G., Jaffe, A. H., Jaffe, T. R., Jones, W. C., Juvela, M., Keihanen, E., Keskitalo, R., Kisner, T. S., Kneissl, R., Knoche, J., Knox, L., Kunz, M., Kurki-Suonio, H., Lagache, G., Lahteenmaki, A., Lamarre, J. M., Lasenby, A., Lattanzi, M., Lawrence, C. R., Leahy, J. P., Leonardi, R., Lesgourgues, J., Levrier, F., Lewis, A., Liguori, M., Lilje, P. B., Linden-Vornle, M., Lopez-Caniego, M., Lubin, P. M., Macias-Perez, J. F., Maggio, G., Maino, D., Mandolesi, N., Mangilli, A., Marchini, A., Martin, P. G., Martinelli, M., Martinez-Gonzalez, E., Masi, S., Matarrese, S., Mazzotta, P., McGehee, P., Meinhold, P. R., Melchiorri, A., Melin, J. B., Mendes, L., Mennella, A., Migliaccio, M., Millea, M., Mitra, S., Miville-Deschenes, M. A., Moneti, A., Montier, L., Morgante, G., Mortlock, D., Moss, A., Munshi, D., Murphy, J. A., Naselsky, P., Nati, F., Natoli, P., Netterfield, C. B., Norgaard-Nielsen, H. U., Noviello, F., Novikov, D., Novikov, I., Oxborrow, C. A., Paci, F., Pagano, L., Pajot, F., Paladini, R., Paoletti, D., Partridge, B., Pasian, F., Patanchon, G., Pearson, T. J., Perdereau, O., Perotto, L., Perrotta, F., Pettorino, V., Piacentini, F., Piat, M., Pierpaoli, E., Pietrobon, D., Plaszczynski, S., Pointecouteau, E., Polenta, G., Popa, L., Pratt, G. W., Prezeau, G., Prunet, S., Puget, J. L., Rachen, J. P., Reach, W. T., Rebolo, R., Reinecke, M., Remazeilles, M., Renault, C., Renzi, A., Ristorcelli, I., Rocha, G., Rosset, C., Rossetti, M., Roudier, G., d'Orfeuil, B. R., Rowan-Robinson, M., Rubino-Martin, J. A., Rusholme, B., Said, N., Salvatelli, V., Salvati, L., Sandri, M., Santos, D., Savelainen, M., Savini, G., Scott, D., Seiffert, M. D., Serra, P., Shellard, E. P. S., Spencer, L. D., Spinelli, M., Stolyarov, V., Stompor, R., Sudiwala, R., Sunyaev, R., Sutton, D., Suur-Uski, A. S., Sygnet, J. F., Tauber, J. A., Terenzi, L., Toffolatti, L., Tomasi, M., Tristram, M., Trombetti, T., Tucci, M., Tuovinen, J., Turler, M., Umana, G., Valenziano, L., Valiviita, J., Van Tent, B., Vielva, P., Villa, F., Wade, L. A., Wandelt, B. D., Wehus, I. K., White, M., White, S. D. M., Wilkinson, A., Yvon, D., Zacchei, A., and Zonca, A. (2015). Planck 2015 results. XIII. Cosmological parameters. *arXiv.org*.

Theuns, T., Bernardi, M., Frieman, J., Hewett, P., Schaye, J., Sheth, R. K., and Subbarao, M. (2002). Detection of He II Reionization in the Sloan Digital Sky Survey Quasar Sample. *The Astrophysical Journal*, 574(2):L111–L114.

Turk, M. J., Smith, B. D., Oishi, J. S., Skory, S., Skillman, S. W., Abel, T., and Norman, M. L. (2011a). yt: A Multi-code Analysis Toolkit for Astrophysical Simulation Data. *The Astrophysical Journal Supplement Series*, 192:9–+.

Turk, M. J., Smith, B. D., Oishi, J. S., Skory, S., Skillman, S. W., Abel, T., and Norman, M. L. (2011b). yt: A Multi-code Analysis Toolkit for Astrophysical

Simulation Data. *The Astrophysical Journal Supplement Series*, 192:9–+.

Tytler, D., Paschos, P., Kirkman, D., Norman, M. L., and Jena, T. (2009). The effect of large-scale power on simulated spectra of the Lyα forest. *Monthly Notices of the Royal Astronomical Society*, 393(3):723–758.

Vanden Berk, D. E., Collaboration, S., and Collaboration, S. (2001). Composite Quasar Spectra From the Sloan Digital Sky Survey. *The Astronomical Journal*, 122(2):549.

Viel, M., Haehnelt, M. G., and Springel, V. (2004). Inferring the dark matter power spectrum from the Lyman α forest in high-resolution QSO absorption spectra. *Monthly Notices of the Royal Astronomical Society*, 354:684–694.

Weinberg, D. H., Davé, R., Katz, N., and Kollmeier, J. A. (2003). The Lyman-α Forest as a Cosmological Tool. In Holt, S. H. and Reynolds, C. S., editors, *The Emergence of Cosmic Structure*, volume 666 of *American Institute of Physics Conference Series*, pages 157–169.

Wise, J. H. and Abel, T. (2011). ENZO+MORAY: radiation hydrodynamics adaptive mesh refinement simulations with adaptive ray tracing. *Monthly Notices of the Royal Astronomical Society*, 414:3458–3491.

Wise, J. H., Demchenko, V. G., Halicek, M. T., Norman, M. L., Turk, M. J., Abel, T., and Smith, B. D. (2014). The birth of a galaxy - III. Propelling reionization with the faintest galaxies. *Monthly Notices of the Royal Astronomical Society*, 442:2560–2579.

Worseck, G., Prochaska, J. X., Hennawi, J. F., and McQuinn, M. (2014). Early and Extended Helium Reionization Over More Than 600 Million Years of Cosmic Time. *arXiv.org*, page 7405.

Worseck, G., Prochaska, J. X., McQuinn, M., Dall'Aglio, A., Fechner, C., Hennawi, J. F., Reimers, D., Richter, P., and Wisotzki, L. (2011). The End of Helium Reionization at z ˜= 2.7 Inferred from Cosmic Variance in HST/COS He II Lyα Absorption Spectra. *The Astrophysical Journal Letters*, 733:L24.

Wyithe, J. S. B. and Loeb, A. (2013). A suppressed contribution of low-mass galaxies to reionization due to supernova feedback. *Monthly Notices of the Royal Astronomical Society*, 428:2741–2754.

Xu, H., Wise, J. H., Norman, M. L., Ahn, K., and O'Shea, B. W. (2016a). Galaxy Properties and UV Escape Fractions During Epoch of Reionization: Results from the Renaissance Simulations. *ArXiv e-prints*.

Xu, H., Wise, J. H., Norman, M. L., Ahn, K., and O'Shea, B. W. (2016b). Galaxy Properties and UV Escape Fractions During Epoch of Reionization: Results from the Renaissance Simulations. *ArXiv e-prints*.

Zaldarriaga, M., Hui, L., and Tegmark, M. (2001). Constraints from the Lyα Forest Power Spectrum. *The Astrophysical Journal*, 557(2):519–526.

Zhang, Y., Meiksin, A., Anninos, P., and Norman, M. L. (1997). Physical Properties of the Lyman Alpha Forest in a Cold Dark Matter Cosmology. *arXiv.org*.

Zheng, W., Kriss, G. A., Deharveng, J. M., Dixon, W. V., Kruk, J. W., Shull, J. M., Giroux, M. L., Morton, D. C., Williger, G. M., Friedman, S. D., and Moos, H. W. (2004). A Study of the Reionization History of Intergalactic Helium with FUSE and the Very Large Telescope. *The Astrophysical Journal*, 605(2):631–644.