INAUGURAL DISSERTATION

zur
Erlangung der Doktorwürde
der
Naturwissenschaftlich-Mathematischen
Gesamtfakultät
der Ruprecht-Karls-Universität
Heidelberg

vorgelegt von

Hector Hiß, MSc
aus
São Paulo, Brasilien

Tag der mündlichen Prüfung: 15.10.2019

Doctoral Thesis

Measuring the thermal evolution of the intergalactic medium

based on the statistical distribution of absorption line parameters in the Lyα forest

Hector Hiß

August 30th, 2019

Universität Heidelberg
Supervisor & First Referee: Prof. Dr. Joseph F. HENNAWI
Second Referee: apl. Prof. Dr. Simon GLOVER

Abstract

The thermal evolution of the low density intergalactic medium (IGM) is a major diagnostic tool for understanding the nature and evolution of the predominant component of baryonic matter in the universe. In this study I present different approaches for measuring the thermal state of the IGM at different ages of the universe, in order to understand how it is affected by reionization processes. The main observable used to probe the thermal state of the gas is the so-called Lyman-α forest. This observable consists of a series of absorption lines in the spectra of distant quasi-stellar objects (QSOs) which arise due to the presence of residual intervening neutral hydrogen in the IGM between the observer and the QSO. Decomposing the Lyman-α forest into discrete absorption profiles allows one to explore how the distribution of Lyman-α absorption line widths and column densities (b-$N_{\mathrm{H\,I}}$ distribution) depends on the thermal state of the gas, which is characterized by a temperature-density relation.

In this thesis, I quantify the parameters of the temperature-density relation using high quality UVES and HIRES QSO spectra and state of the art cosmological hydrodynamic simulations. In the first part of this study, I apply a traditional cutoff fitting method to the b-$N_{\mathrm{H\,I}}$ distribution of the QSO spectra. Using simulations, I calibrate how the position of the cutoff in the b-$N_{\mathrm{H\,I}}$ distribution relates to the thermal state of the IGM. I find that the thermal evolution of the IGM shows clear signatures of He II reionization at $2 \leq z \leq 3.4$. In the second part of this thesis, I present a novel statistical method for constraining the thermal state of the IGM using the full shape of the b-$N_{\mathrm{H\,I}}$ distribution. I show that this method is more accurate and precise than the traditional cutoff fitting approach, by applying it to mock data realizations. I confirm this by applying it to observational data at $z = 2$. Finally, using this novel method, I quantify for the first time the parameters of the temperature-density relation at low redshift ($z = 0.1$) using the b-$N_{\mathrm{H\,I}}$ distribution, and find broad agreement with theoretical expectations.

Overall, this thesis demonstrates that the b-$N_{\mathrm{H\,I}}$ distribution is a powerful statistical tool for studying the intergalactic medium and can place strong constraints on the evolution of its thermal state.

Zusammenfassung

Die thermische Entwicklung des diffusen intergalaktischen Mediums (IGM) ist einer der wichtigsten Indikatoren, um die Natur und Entwicklung des größten Teils an baryonischer Materie im Universum zu diagnostizieren. In dieser Arbeit stelle ich verschiedene Methoden vor, um die thermische Entwicklung des IGM zu messen und seine Reaktion auf Reionisationsphasen zu verstehen. Die wichtigste Beobachtungsgröße, die Informationen über den thermischen Zustand des IGM trägt, ist der sogenannte Lyman-α Wald. Dieser besteht aus mehreren Absorptionslinien in den Spektren von weit entfernten Quasaren (QSO), die aufgrund von verbliebenem neutralen Wasserstoff zwischen dem Beobachter und den Quasaren entstehen. Die Zerlegung des Lyman-α Waldes in einzelnen Absorptionsprofile ermöglicht die Analyse der Verteilung des Doppler-Verbreiterungsparameters und der Säulendichte (b-$N_{\mathrm{H\,I}}$-Verteilung) sowie wie diese im Zusammenhang zum thermischen Zustand des IGM stehen. Dieser wird durch ein wohldefiniertes Temperatur-Dichte-Verhältnis beschrieben.

In dieser Arbeite quantifiziere ich die Parameter des Temperatur-Dichte-Verhältnisses mit Hilfe von hochwertigen UVES und HIRES QSO Spektren in Kombination mit modernen hydrodynamischen Simulationen. Im ersten Teil wird eine traditionelle Methode angewandt, bei welcher eine Verteilungsrandlinie an die b-$N_{\mathrm{H\,I}}$ Verteilung angepasst wird. Mit Hilfe von Simulationen wird der Zusammenhang zwischen der Position der Verteilungsrandlinie und den thermischen Parametern kalibriert. Ich stelle fest, dass die thermische Entwicklung des IGM im Rotverschiebungsintervall $2 \leq z \leq 3.4$ klare Anzeichen von He II Reionisation zeigt. Im zweiten Teil präsentiere ich eine neue statistische Methode, um den thermischen Zustand des IGM anhand der vollständigen Form der b-$N_{\mathrm{H\,I}}$-Verteilung zu messen. Ich zeige, dass diese Methode präziser als die Methode der Verteilungsrandlinien-Anpassung ist und über eine höhere Richtigkeit verfügt. Das zeige ich, indem ich mehrere Messungen mit simulierten Daten durchführe. Ich bestätige die Genauigkeit dieser neuen Methode, indem ich sie auf $z = 2$ Daten anwende. Weiterhin wird diese neue Methode benutzt, um zum ersten Mal den thermischen Zustand des IGM bei niedriger Rotverschiebung ($z = 0.1$) aus der b-$N_{\mathrm{H\,I}}$-Verteilung zu messen. Das damit erhaltene Resultat ist in gutem Einklang mit theoretischen Erwartungen.

Die vorliegende Arbeit zeigt, dass die b-$N_{\mathrm{H\,I}}$-Verteilung ein mächtiges statistisches Mittel ist, um den thermischen Zustand des intergalaktischen Gases und dessen Entwicklung zu entschlüsseln.

Preface

Some days I wake up
Thinking and wanting to know
From whence comes
Our impulse
To probe space

- Jorge Ben, *Errare Humanum Est* 1974

Often as a scientist one is faced with the question, mostly by a curious friend, "What is your research about exactly"? To which I typically answer "I try to understand the properties of the very diffuse gas that lies between galaxies". Normally, I will add a simple explanation of what galaxies are and how gigantic the space between them is. Knowing the efforts and resources involved in scientific research, I always fear that the next question will be "Why would you do that?". A question to which I personally do not have a better short answer to than "because why not" or "because it could have a return to society... eventually". To my surprise however, most of the time the person I am talking to will nod and ask "*How* is it even possible to see this gas?".

This reaction makes me happy for different reasons: for one, all of a sudden the blood, sweat and tears that I and many others invested into understanding this component of the universe does not appear as absurd as I might have initially feared. The listener seems to have some sympathy for our struggles, which is in itself remarkable. And secondly, because I know that the method I use to study this gas is actually very intuitive and that the listener may be surprised by the simplicity of it.

The typical answer I give is as follows: Imagine there is a very distant object that is not in our galaxy and that it is so bright that it is visible from Earth even though it is incredibly distant. Imagine there is absolutely nothing between us and this bright light. From Earth we could observe this object by simply collecting its light. Imagine now that we add this intergalactic gas between us and the bright object. Now, as light travels to us it will interact with this gas and whatever light arrives at the observer will be affected by the presence of the intervening medium. If one has knowledge about how the object is *supposed to shine* and has observed how it *is shining* after passing through this gas, then it is possible to learn about the gas that is responsible for this difference. This allows us to study the major reservoir of baryonic matter in the universe and the fuel for the formation of galaxies, the intergalactic medium.

Suddenly a new world opens up for the listener. We all have been exposed to beautiful imagery of the night-sky consisting of majestic bright objects such as stars, galaxies and colorful nebulae, but the general listener might never have been exposed to the idea of using bright objects to study what lies between us and it. Exactly in these moments, when I see the realization in the face of others is when I mostly appreciate the work that I am allowed to be part of. And often it takes a good question to put things in perspective and to learn to appreciate the beauty and tangibility of one's research.

The kind of technique I described in simple terms above is by no means new. Kirchoff's third law of spectroscopy states that an absorption line spectrum is seen when a source of a continuous spectrum is viewed behind a cool gas under pressure. This approach has found use in many fields of physics and chemistry and beyond that in biology, medicine and even forensics. Absorption spectroscopy in astronomy and cosmology has proven itself as an invaluable tool, especially in light of the great improvements in modern instrumentation that allow us to infer the detailed physical characteristics of the gas responsible for the absorption. The other essential piece of the puzzle that has seen major improvements are the techniques that allow one to simulate the evolution of the intergalactic gas in the universe.

The objective of this thesis is to combine information gathered from the observed universe (as seen through the light of absorption profiles) and compare it to the output of simulated universes where we have control over the global parameters defining the physical characteristics of the intergalactic gas. This approach allows us to quantify what scenarios are more likely to be driving the evolution of this gas, i.e. what parametrization of the thermal state of the gas is preferred given observed data. In detail, the questions I wish to address are: do we see evidence for heating events during the cosmic evolution of intergalactic matter and if yes, what are the sources driving it? Can we reproduce past observations of the thermal evolution of the intergalactic medium when using classic methods combined with new and more abundant datasets and simulations? Can we go beyond the classic methods and develop a new approach that provides better constrains and is less prone to systematic effects?

This thesis presents an endeavor to find answers to some of these questions. During the course of my PhD I used absorption line catalogs generated from state of the art simulations and observed data in order to:

- Quantify two billion years of evolution of the thermal state of the intergalactic gas
- Develop, test and apply a novel statistical method for carrying out thermal state measurements in the intergalactic medium

Ultimately, the goal of my field of study is to refine our understanding of the thermal evolution of baryonic matter in the universe and the sources that are driving it. So, if a curious person should ask "Why would you do that?", I guess maybe I could say: "We are curious to know how exactly things around us came to be". In other words, understanding the history of the fuel of galaxy formation is essential for understanding where we come from.

July 29th 2019, Hector Hiss

Contents

1. Introduction

> *Accept the universe*
> *As the gods gave it to you.*
> *If the gods wanted to give you something else*
> *They'd have done it.*
>
> *If there are other matters and other worlds*
> *Then there are*
>
> - Alberto Caeiro, *Aceita o Universo* 1917

1.1. Context

The intergalactic medium (IGM) is the low density reservoir of baryonic matter that serves as fuel for the formation of galaxies. The evolution of the thermal state of the diffuse IGM provides us with insight into the nature and evolution of the bulk ($\gtrsim 90\%$ at $z > 2$) of baryonic matter in the universe (Meiksin 2009; McQuinn & Upton Sanderbeck 2016; de Graaff et al. 2019; Tanimura et al. 2019). In order to understand the history of this gas' thermal state, it is important to understand all the different processes that can affect it during the evolution of the universe.

In the current cosmological paradigm all baryonic matter was initially highly ionized. It is not until 350.000 years after the Big Bang (at redshift $z \sim 1100$) that the universe expanded and thus cooled down enough in order to allow free electrons to bind to protons for the first time, resulting in the first neutral hydrogen atoms in the universe. Before that, the mean free path of photons in the cosmic plasma was very short, due to the high density of free electrons. Thus this transition of the universe to bound electrons allowed photons to travel freely for the first time, i.e. the universe became *transparent*. This major phase transition of baryonic matter (predominantly hydrogen and helium) is known as recombination and it is characterized by the emission of a highly homogeneous global radiation field that we now observe ubiquitously in the micro-wave regime (Alpher & Herman 1948). This so-called Cosmic Microwave Background (CMB) is one of the most important cosmological probes, as small amplitude fluctuations within this background proved to be very sensitive to the parameters driving cosmological expansion (Smoot et al. 1992; Hinshaw et al. 2013; Planck Collaboration et al. 2014).

After recombination, the baryonic matter stayed neutral for about another billion years (down to $z \sim 6$ through the "dark ages", see Pritchard & Loeb (2012) for a review). During the dark ages, this pristine gas was able to cool and collapse into hierarchically growing dark matter (DM) halos, which culminated in the formation of the first galaxies, consisting of the first population III stars (Faucher-Giguère et al. 2008a; Robertson et al. 2015). These galaxies are in all likelihood the main contributors powering another global phase transition of baryonic matter in the universe (Bouwens et al. 2015; Finkelstein et al. 2015), the so-called *reionization* (or hydrogen reionization) phase. Before reionization, all baryonic matter was in a neutral state and there were no sources of photons that could separate the electrons from their nuclei. This picture changed when the first galaxies formed.

Galactic emission of photons with sufficient energy to ionize hydrogen $E_{\mathrm{H\,I}\rightarrow\mathrm{H\,II}} \gtrsim 13.6$ eV is expected to have driven hydrogen reionization (H I→H II)[1] which is a process that is believed to be completed by redshift $z \sim 6$ (Fan et al. 2006; McGreer et al. 2015; Becker et al. 2015; Bosman et al. 2018; Eilers et al. 2018) (for reviews about the hydrogen reionization epoch see Loeb & Barkana (2001); Mesinger (2016)). Recent measurements based on damped absorption in the spectra of the highest redshift spectrum also indicate that the IGM at $z \sim 7$ is mostly neutral (Bañados et al. 2018; Davies et al. 2018). When a hydrogen atom is ionized by a photon with energy higher than the ionization threshold, the energy difference will be deposited into the photoelectron in the form of kinetic energy. This will result in a net heating of the gas. During this process the temperature of this gas is expected to increase from a few Kelvin to a temperature on the order of 10 000 K. Spatially, hydrogen reionization is expected to be an inside-out process, meaning that individual galaxies will first ionize their surroundings, such that reionization will happen in a *patchy* way. Eventually these ionized regions will overlap, giving rise to an ultraviolet background (UVB), an integrated global ionizing field of photons that are not locally absorbed by the neutral gas (Haardt & Madau 2012).

During the next 2 billion years, the population of very luminous active galactic nuclei (AGN) galaxies, i.e. quasi-stellar objects (QSO), builds up. This means that during this time high energy photons released by AGN become increasingly available to power a second phase transition (Furlanetto & Oh 2008; McQuinn et al. 2009), namely the second reionization of helium , He II→He III with a higher ionizing threshold of $E \gtrsim E_{\mathrm{He\,II}\rightarrow\mathrm{He\,III}} = 54.4$ eV (see e.g. Madau & Meiksin 1994; Miralda-Escudé et al. 2000; Dixon & Furlanetto 2009; Compostella et al. 2013, 2014; Syphers & Shull 2014; Dixon

---

[1] Due to comparable ionization thresholds, and $E_{\mathrm{He\,I}\rightarrow\mathrm{He\,II}} \gtrsim 24.6$ eV, it is normally assumed that helium is singly ionized (He I→He II) along with H I.

et al. 2014).[2] Due to the requirement of high-energy photons from this QSO population, this process becomes only possible at much later times and is expected to be completed by $z \simeq 2.7$ (see e.g. Worseck et al. 2011, 2018). During this second reionization phase, the temperature of the IGM is expected to increase by another $5000\,\mathrm{K}$.

The key aspect for this work is that these reionization epochs are expected to significantly alter the thermal history of the IGM. Furthermore, the exact timing and evolution of these reionization processes is dependent on the properties of the gas and the sources responsible for powering them. Ultimately, understanding the detailed evolution of the temperature of the IGM will allow us to test our models and simulations and enhance our understanding of the history of the baryonic component of the universe.

In this thesis I will present different approaches for measuring the temperature of the IGM at different ages of the universe. This chapter will focus on the introduction of the parameters that govern the thermal state of the IGM and how they are affected by reionization processes. The main observable that is used to probe the thermal state of the gas, the so-called Lyman-α (Lyα) forest, and the method by which we decompose it into individual absorption profiles from which we can infer temperatures, will be introduced in the remainder of this chapter.

In chapter 2 I will apply a classic method of inferring the temperature of the IGM from absorption profiles to a large number of observed spectra and state of the art cosmological simulations. This method will be used to measure the temperature evolution in the redshift range $2 \leq z \leq 3.4$ (covering the time of He II reionization). In chapter 3 I will develop, test and apply a new method at $z = 2$ that is statistically more powerful and less prone to systematic errors for constraining temperatures using the same observable. In chapter 4 I will explore the possibility of applying this new approach to low redshift data ($z = 0.1$). Finally I will conclude this thesis in chapter 5.

1.2. Reionization and Thermal Evolution of the IGM

1.2.1. The Temperature-Density Relation in the IGM

The thermal state of the IGM is mainly dictated by three major factors[3]. The major source of heating is the aforementioned UVB, while the adiabatic expansion of the universe and Compton cooling of the gas are the major sources of cooling. The combined effects of photoionization heating and cooling effects result in a tight power law temperature-density relation (TDR) (Hui & Gnedin 1997; McQuinn & Upton Sanderbeck

---

[2] Note that current discussions exist considering that early QSO might have played a role in the process of hydrogen reionization (Madau & Haardt 2015; Khaire et al. 2016)

[3] This is true when the major photoionization source is the UVB, i.e. long after reionization effects where local ionizing sources are driving photoionization

2016) of the form:

$$T(\rho) = T_0 \left(\frac{\rho}{\rho_0}\right)^{\gamma-1} =: T_0 \Delta^{\gamma-1}, \tag{1.1}$$

where T_0 is the temperature at mean baryonic density ρ_0 and the power law index γ sets the temperature difference between overdensities ($\Delta > 1$) and underdensities ($\Delta < 1$). It can be shown that the index γ is expected to approach ~ 1.7 asymptotically if no major reionization processes are taking place (Hui & Gnedin 1997; McQuinn & Upton Sanderbeck 2016).

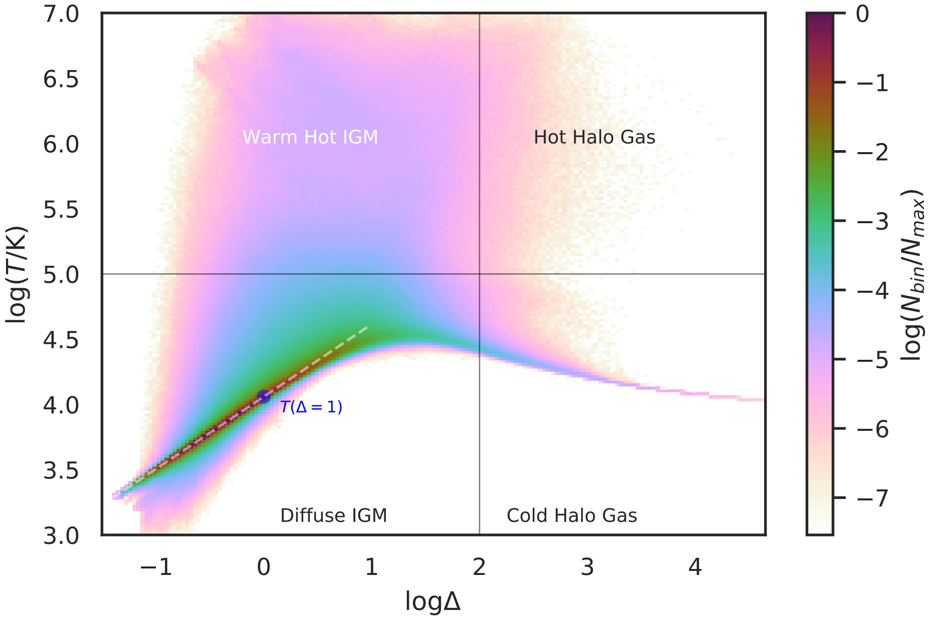


Figure 1.1. Temperature-density distribution from a Nyx hydrodynamic simulation run (Almgren et al. 2013; Lukić et al. 2015) which will be introduced in § 2.2. The distribution of temperatures and densities in a snapshot at $z = 3$ is color-coded using the logarithm of the number of cells within a bin, i.e. the volume weighted probability distribution. The majority of the intergalactic gas (with overdensities around the mean $\log\Delta = 0$) lies on a well defined power law relation (Hui & Gnedin 1997; McQuinn & Upton Sanderbeck 2016). The mean density ($\log\Delta = 0$) is indicated by a blue circle. A TDR (eqn. 1.1) with $T_0 = 11500\,\mathrm{K}$ and $\gamma = 1.54$ is shown as a white dashed line. Gas may be heated up to temperatures greater than 10^5 K when mechanisms other than photoionization and adiabatic expansion, such as shock heating, are taking place (Warm Hot IGM or WHIM). Overdense cold halo gas at $\log\Delta \gtrsim 2$ is able to cool efficiently and is not well described by the relation. The top right portion of the distribution represents the hot halo gas that in in the process of virialization.

An intuitive way of understanding the density dependence of the temperature is by picturing that denser gas has a higher optical depth and will more likely experience heating. Also, high density clouds of gas are less coupled to expansion and therefore experience less cooling. Since the majority of baryonic matter is in the IGM, for most of the cosmic history, the majority of baryons is well described by this relation.

Note that although the TDR is expected to have positive $(\gamma - 1)$ values in the standard scenario, the index γ allows for an isothermal IGM when $\gamma = 1$, i.e. all gas has a constant temperature independent of density. Inverted TDRs, where underdensities have higher temperatures than overdensities, are described by values of $\gamma < 1$. Although the TDR favors positive values of γ in the standard photo-heating vs expansion scenario, there are other postulated heating mechanisms such as *Blazar Heating* that are capable of inverting the TDR (see Puchwein et al. 2012; Pfrommer et al. 2013; Lamberts et al. 2015) through electron-positron pair production from very high energy photons.

Figure 1.1 illustrates the thermal state of the IGM by showing the distribution of the temperature and the density of gas cells in a snapshot of a Nyx cosmological hydrodynamic simulation (see § 3.1) that includes all necessary cooling and heating mechanisms. The majority of the low density gas with overdensities $\Delta < 10$ is well described by the TDR shown as a white dashed line. This gas corresponds to very low densities and is expected to be far from collapsing into galaxies.

Figure 1.1 also roughly illustrates the other three phases of the IGM that are not well described by a global TDR (Lukić et al. 2015):

1. *Warm hot IGM (WHIM)*: a fraction of the rarefied gas is heated up to temperatures higher than 10^5 K due to accretion driven shock heating. This phase is specially present at low redshift ($z \lesssim 1$) when underdensities are increasingly fed to overdensities.

2. *Cold halo or condensed gas*: collapsing high density ($\Delta \gtrsim 100$) gas that is able to cool efficiently.

3. *Hot halo gas*: high density gas in the process of virialization, i.e. systems of gravitationally interacting gas that are not stable.

1.2.2. The Effect of Cosmic Reionization Processes

The intergalactic gas is believed to have undergone two major reionization processes, namely the reionization of H I and He I, followed later by the reionization of He II. The TDR can be used as a diagnostic tool for understanding these reionization phase transitions. This is possible, because the photoionization heating during these phases will affect the thermal state of the gas which is described by the TDR.

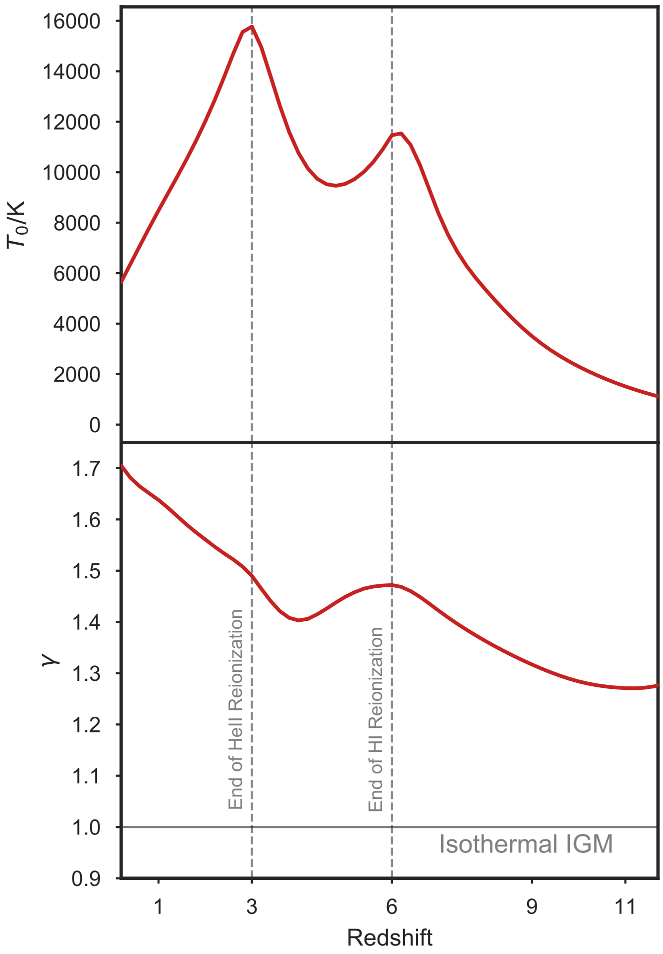


Figure 1.2. Evolution the TDR based on the output of a Nyx cosmological hydrodynamic simulation. **Top panel:** The evolution shows how T_0 increases during the events of H I and He I reionization (from $z \sim 12$ to 8.). The gas proceeds to cool down slightly from $z \sim 6$ to $z \sim 4$ due to expansion until the increasing QSO population initiates He II reionization. This second reionization process is responsible for the second peak in T_0 at $z \sim 3$. **Bottom panel:** The power law index γ also responds to global reionization processes. Heating the IGM at all densities results in a flattening of the TDR, i.e. it approaches an isothermal state during H I and He I reionization.

For illustration, the evolution of the TDR parameters for a simulated reionization model is shown in Figure 1.2. In essence, gas moves into hotter (higher T_0) and flatter (lower γ) TDRs as reionization processes proceed. The first temperature peak, with $T_0 \simeq 10\,000$ K, occurs at $z \sim 6$ due to the photoheating of the IGM during H I reionization. When H I reionization is complete, cosmic expansion causes the gas to cool until He II

reionization initiates a new heating process, which peaks at $z \sim 3$ with $T_0 \simeq 15\,000$ K. During this heating phase, i.e. before T_0 peaks, γ stays below the asymptotic value of 1.7, which is expected when the gas is heated at all densities. After He II reionization is complete ($z < 3$), the IGM is expected to cool down adiabatically and γ will approach its asymptotic value of 1.7.

Here we should note that, in reality, reionization can not be treated as a homogeneous and instantaneous process. This means that the gas will not leave the pre-reionization TDR all at once everywhere, but rather with some delay and spatial dependency. These effects will inevitably affect the temperature-density distribution of the gas by adding scatter around the TDR[4] or even by creating multivalued relations (McQuinn et al. 2009; Compostella et al. 2013). Although these effects play a role during reionization, analytic calculations (McQuinn & Upton Sanderbeck 2016) suggest that the gas is expected to relax back into the tight TDR from eqn. 1.1 after several hundred Myr.

Another assumption frequently used in reionization studies is that of photoionization equilibrium. While true at most times in the IGM's thermal history, this assumption might break down amidst an epoch of intense thermal heating, as the IGM gas does not react instantaneously to sudden changes in the thermal state. However, this effect is also expected to play a minor role long after reionization epochs (Puchwein et al. 2012, 2019).

1.3. The Lyman-α Forest

We have reviewed the thermal evolution of the IGM and how it is sensitive to reionization processes. Unfortunately, there is no way of measuring the temperature or density of the intergalactic gas directly. Thus, we have to rely on other observables that are sensitive to the TDR, such as the distribution of absorption line shapes in the so-called Lyα forest. In this section we will introduce the Lyα forest, while the technique used will be discussed in detail in § 1.5.

The diffuse nature of the IGM renders emission lines difficult to observe, so one must rely on the absorption features caused by the IGM in order to gain insights into its properties. Although predominantly photoionized, residual neutral hydrogen in the diffuse IGM gives rise to Lyα[5] absorption. This series of Lyα absorption profiles is called Lyα forest and has been established as the premier probe of the IGM and cosmic structure at redshifts $z \lesssim 6$. The Lyα forest was first identified by Lynds (1971) based on observed hydrogen absorption towards the QSO 3C 9 (Gunn & Peterson 1965) and is a

---

[4] These fluctuations can be due to temperature (D'Aloisio et al. 2015) and/or UVB fluctuations (Davies et al. 2017)

[5] Lyα refers to the hydrogen transition from the ground state $n = 1$ to $n = 2$, with n being the principal quantum number

spectral feature generally observed in the spectra of QSOs. It consists of a series of Lyα absorption lines that lie at shorter wavelengths (blueward) of the Lyα emission spike of a QSO. The most interesting property of these lines is that they are not intrinsic to the QSO spectra, but originate in the resonant absorption of photons by the intervening residual H I in the IGM. In this thesis I will explore the thermal evolution of the IGM through its effect on the shape of Lyα absorption lines in QSO spectra.

Lyα absorption comes about when a photon with energy of $E_{\mathrm{Ly}\alpha} \simeq 10.2\,\mathrm{eV}$ (corresponding to $\lambda_{\mathrm{Ly}\alpha} \sim 1215.67$ Å) released by the QSO excites the electron in a neutral hydrogen atom in the IGM from the ground-state into the first excited state. In a static universe, this would mean that all Lyα absorption by the intervening gas would fall at $\lambda_{\mathrm{Ly}\alpha} = 1215.67$, but in the context of an expanding universe the QSO spectrum is redshifted as its light travels through the IGM. Due to cosmic expansion, the QSO emission at a given wavelength $\lambda_{\mathrm{Q,em}}$ emitted at redshift z_{Q} will be shifted toward a longer wavelength when it arrives at the observer:

$$\lambda_{\mathrm{Q,obs}} = \lambda_{\mathrm{Q,em}}(1 + z_{\mathrm{Q}}), \tag{1.2}$$

where $\lambda_{\mathrm{Q,obs}}$ is the wavelength as perceived by the observer.

This redshifting of the QSO's spectrum due to cosmic expansion happens gradually as light is traveling to the observer. This implies that the absorption wavelength for Lyα will be shifted to different positions in the QSO continuum as the light travels. Intervening gas will simply absorb photons at its rest-frame Lyα transition wavelength, resulting in absorption profiles in the QSO continuum at different wavelengths, depending on how long after emission it took for the absorption to take place. Note that the absorption signal is caused by neutral hydrogen in a predominantly ionized IGM (at $z < 6$), i.e. a neutral fraction of the order of $n_{\mathrm{H\,I}} \sim 10^{-5}$ is responsible for this absorption. Luckily the Lyα transition has a high oscillator strength (see § 1.4.1) and will absorb strongly despite the low neutral fraction.

A graphic depiction of the origin of the Lyα forest is shown in Figure 1.3. A QSO emits its continuum plus a Lyα emission line (labeled "a", upper panel) which are perceived later by an observer (in an expanding universe). There are three Lyα absorbing clouds (labeled "b", "c" and "d") between the observer and the QSO. At the timestep $t_1 = 0$ we see the QSO spectrum as observed at its position. It consists of a continuum with a Lyα emission peak of the QSO at $\lambda \simeq 1216$. As light is traveling from the QSO to cloud "b", the universe expands causing the QSO continuum to be redshifted. At the timestep $t_2 > 0$ we see that once the QSO light arrived at cloud "b", the Lyα absorption wavelength does not match that of the QSO emission peak anymore. Due to the QSOs redshifting, cloud "b" will absorb at a shorter wavelength relative to the QSO emission. Analogously, in timestep $t_3 > t_2$, due to further redshifting, both the QSO emission and the absorption feature left by cloud "b" will move toward longer wavelengths. Cloud

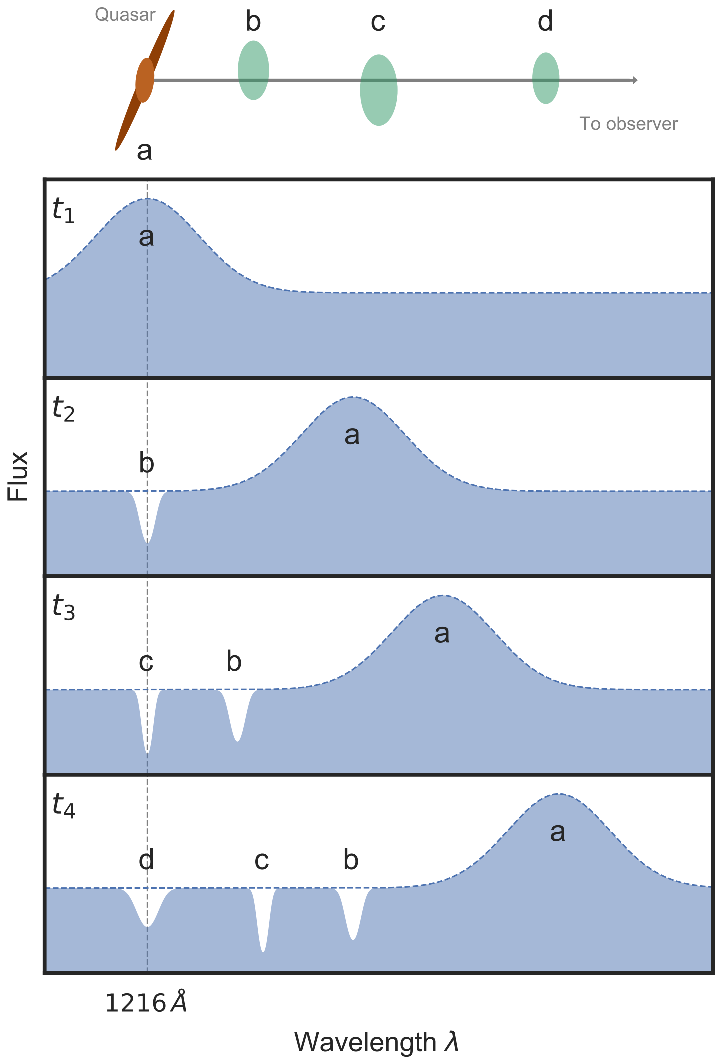


Figure 1.3. Simplified sketch of how the Lyα forest originates. As light from a QSO travels to the observer, residual clouds of neutral hydrogen result in absorption profiles. The fact that the universe is expanding during the time the light is traveling causes the QSO continuum to be redshifted and the clouds to absorb at different positions in wavelength relative to the QSO emission peak.

"c" will absorb at a wavelength that is shorter relative to "b" and even shorter relative to the QSO emission peak. At the timestep $t_4 > t_3$ we observe the same effect for absorption for cloud "d". With the above picture in mind, the Lyα forest is essentially a one-dimensional map of the neutral hydrogen distribution between observer and the

QSO. It is worth noting that at high redshift ($z > 5$), the hydrogen neutral fraction in the IGM is high enough to completely absorb the QSO continuum (so-called "Gunn-Peterson troughs"). Therefore this picture is a good description of the Lyα forest for low redshift.

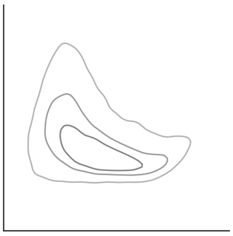
We should keep in mind that this framework is a simplification of the process and that there are other effects that distort this picture. One example is that the absorbing clouds might not be at rest, i.e. they have a net peculiar velocity. This results in an additional shifting of the absorption position in redshift space (Weinberg et al. 2003) toward shorter or longer wavelengths depending on the direction of the line-of-sight component of the peculiar velocity. Such an effect can also cause lines to overlap. Peculiar velocity effects make a one-to-one relation between gas density and underlying DM density fields rather challenging to measure using the Lyα forest.

In Figure 1.4, I present an actual observed QSO spectrum at redshift $z \simeq 2.89$. The expected QSO continuum is shown in red and the observed spectrum is shown in black. There are a number of features in this spectrum that are not explained by our simple picture from before. First, one might notice that there are absorption lines that fall at longer wavelengths than the QSO Lyα emission peak. In the example of the absorption features at 4850 (right of the QSO Lyα emission peak), if these lines would appear redshifted due to peculiar velocities, it would require a Lyα absorber to have line-of-sight velocity away from the observer higher than 40×10^3 km s^{-1}in order to absorb at these wavelengths. Therefore, these absorbers can not be due to hydrogen. In fact, these features are caused by intervening metals present in the IGM that absorb at different wavelengths. These are often associated with high density hydrogen absorbers, presumably star forming galaxies, but might also be seen in the low density IGM itself. Metals can appear anywhere in the spectrum (also within the Lyα forest), and have to be accounted for so they are not mistaken with real Lyα absorption. Later in § 2.1.4 I will discuss in detail how we remove this sort of contamination from our dataset.

Another caveat that should be considered is the fact that there are higher order hydrogen transitions (indicated in Figure 1.4 with vertical lines) associated with the absorbers, namely the Lyβ (rest frame 1025 Å), Lyγ and higher order transitions. Naturally, one should expect that each of these transitions would result in a particular *forest* and that they should eventually overlap after Lyβ [6]. In this thesis I will focus on the part between Lyα and Lyβ which should have no other hydrogen transitions in it than Lyα.

---

[6] Imagine an absorber at $z = 2$: it will absorb Lyα at 3648 Å and Lyβ at 3075 Å. If the same line of sight encounters another hydrogen absorber at $z \simeq 1.53$ there will be a Lyα absorption feature at 3075 Å and Lyβ at 2593 Å. The wavelengths of Lyα and Lyβ absorption can thus overlap, but, assuming peculiar velocities are small, Lyβ cannot be redshifted into the Lyα forest region.



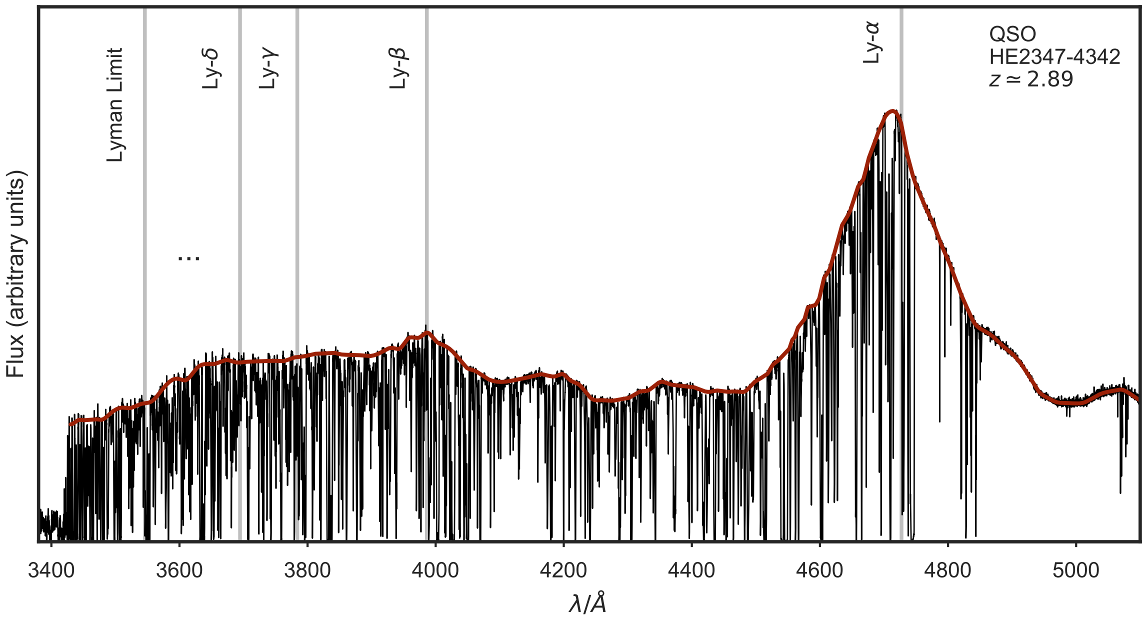


Figure 1.4. The Lyα forest in the sightline of the QSO HE2347-4342 at $z \simeq 2.89$. The broad Lyα emission peak of the QSO is redshifted as described in eqn 1.2 to a wavelength of 4730 Å. The absorption lines at shorter wavelengths than the Lyα peak are caused by the intervening residual neutral hydrogen in the line of sight. The Lyα forest is a result of redshifting of the observed Lyα wavelength with respect to the QSO spectrum as seen by the absorbers in the IGM. H I absorption systems generate a series of absorption profiles of higher order Lyman transitions. Each transition will have a corresponding forest, and different forests can overlap. Metal line absorption can be present throughout the whole spectrum, but is easily visible at wavelengths larger than the QSO's Lyα emission peak. At lower wavelengths than 3410 Å there is complete absorption of Lyman limit photons by an intervening high density absorber. We will discuss how this comes about in § 1.4.4.

1.3.1. Thermal Sensitivity of the Lyα forest

As introduced previously, the Lyα forest is a distorted non-linear probe of the distribution of gas along the line of sight in the intervening IGM. In addition, and especially important in this work, the Lyα forest also carries information about the thermal state of the IGM. This will be discussed in more detail in § 1.4.2, but it should be noted that in general, the shape of the absorption profiles is influenced by the random thermal motions of the gas.

In the literature, different approaches have been used for measuring the parameters of the TDR of the IGM using the Lyα forest. These can be categorized in two groups:

1. Statistical measures of the smoothness of the Lyα forest transmission. This includes various statistical measures such as the power spectrum of the transmitted flux (e.g. Theuns et al. 2000; Zaldarriaga et al. 2001; McDonald et al. 2006; Walther et al. 2018; Khaire et al. 2019; Walther et al. 2019; Boera et al. 2018), the curvature of the smoothed flux field (Becker et al. 2011; Boera et al. 2014), the flux probability distribution function (e.g. Bolton et al. 2008; Viel et al. 2009; Lee et al. 2015), as well as wavelet decompositions of the forest (e.g. Theuns et al. 2002; Lidz et al. 2010; Garzilli et al. 2012).

2. Methods that treat the Lyα forest as a superposition of multiple discrete absorption profiles (Schaye et al. 1999; Ricotti et al. 2000; McDonald et al. 2001; Rudie et al. 2012a; Bolton et al. 2014; Hiss et al. 2018, 2019).

In this work I will study the Lyα forest using the second method, i.e. as a superposition of individual absorption profiles in order to infer the thermal state of the IGM at different redshifts. More specifically, I will fit absorption lines to the Lyα forest of observational and simulated data via *Voigt profiles*. An introduction to Voigt profiles and their sensitivity to the temperature of the absorbing gas will follow in the next section.

1.4. Voigt Profiles and Transmission in the IGM

The aim of this study is to constrain the evolution of the parameters that govern the thermal state of the IGM using the Lyα forest. In order to do so, I will apply a technique that is based on decomposing the Lyα forest into individual absorption (Voigt) profiles.

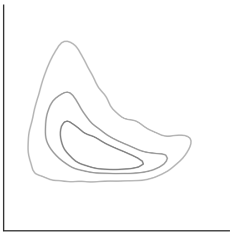
This approach was presented by Haehnelt & Steinmetz (1998) and developed by Schaye et al. (2000); Bryan & Machacek (2000); Ricotti et al. (2000); McDonald et al. (2001) for constraining thermal parameters in the early 2000's. A decade later, in light of an increasing number of high resolution and signal-to-noise ratio (SNR) QSO observations, as well as advancements in hydrodynamic simulations, this method was revisited in studies by Rudie et al. (2012a); Bolton et al. (2014).

The shape of an absorption line is given by the convolution of a Gaussian profile with an Lorentzian profile, i.e.

$$V(x) = \int G(y)L(x-y)\mathrm{d}y, \tag{1.3}$$

where

$$G(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-x^2}{2\sigma^2}\right) \tag{1.4}$$



is a Gaussian function with standard deviation σ and

$$L(x) = \frac{\text{ɣ}}{\pi(x^2 + \text{ɣ}^2)} \tag{1.5}$$

is a Lorentzian profile with scale-parameter ɣ (not to be confused with the power law index γ of the TDR in eqn. 1.1).

Depending on the contribution of the two measures of dispersion σ and γ, the convolution described in eqn. 1.3 will result in three characteristic shapes: **i)** $\gamma \simeq \sigma$: the resulting Voigt profile has a Gaussian core with Lorentzian wings; **ii)** $\gamma \ll \sigma$: the resulting Voigt profile resembles the Gaussian curve; **iii)** $\gamma \gg \sigma$: the resulting Voigt profile resembles the Lorentzian curve. The first two cases are illustrated in Figure 1.5.

1.4.1. The Intrinsic and Thermal Line Profiles

The Voigt profile serves as a good physically motivated shape for absorption profile in our context. This is so, because the intrinsic absorption line shape given by the scattering cross-section of the Lyα transition is approximately Lorentzian. Further, it can be safely assumed that random thermal motions within the absorbing clouds follow a Maxwell-Boltzmann velocity distribution that has a Gaussian shape. This will be discussed in detail for the example of the Lyα transition in the following paragraphs.

The Thermal Line Profile

The hydrogen atoms that give rise to the Lyα forest are not at rest during the absorption. The motion of the atoms with respect to the incoming photons will shift the energy needed to excite their electrons. This process is equivalent to a Doppler shift of the photon's frequency (and therefore energy) from the perspective of the hydrogen atom due to the relative motion.

The distribution of velocities p_v of particles in an ideal gas in thermal equilibrium can be described by the (Gaussian shaped) Maxwell distribution:

$$p_v = \frac{1}{\sqrt{2\pi}\sigma_v} \exp\left(\frac{-(v - v_0)^2}{2\sigma_v^2}\right), \tag{1.6}$$

where v_0 is the mean velocity of the gas particles and the standard deviation can be written as $\sigma_v = \sqrt{(k_B T/m)}$ for pure thermal motion, where m is the particle mass, T the temperature and k_B the Boltzmann constant. Note that only line-of-sight velocities will be important in terms of absorption profile broadening, so we are treating the velocity distribution in one dimension. Throughout this thesis I will use a slightly different

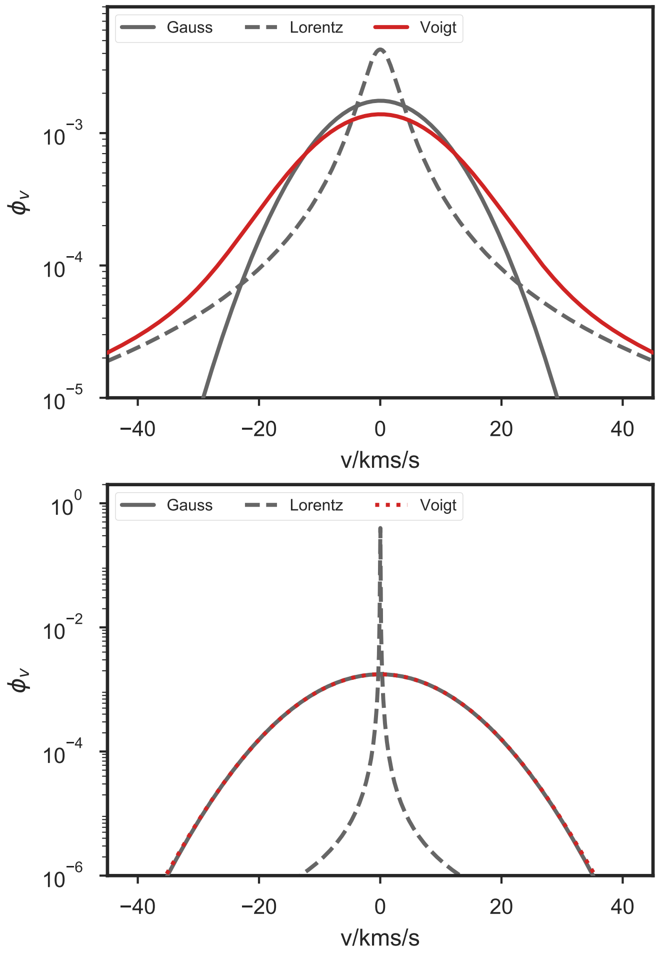
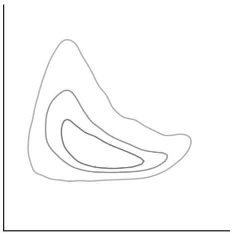


Figure 1.5. Convolutions of Lorentzian and Gaussian profiles and the resulting Voigt profiles. **Top**: A Gaussian profile with Doppler parameter $b = 13\ \mathrm{km\,s^{-1}}$ and a Lorentzian profile with scale parameter $\gamma = 3\ \mathrm{km\,s^{-1}}$. If b and γ are comparable, then the resulting Voigt profile has Lorentz wings but a Gaussian core. **Bottom**: A Lorentz profile with γ based on the intrinsic profile of the Lyα transition (eqn. 1.11). This illustrates the approximation used in eqn 1.13, i.e. a Lorentz profile with $\gamma \ll b$ is approximated by a δ-function. The resulting Voigt profile strongly resembles the Gaussian profile.

convention for the broadening of this distribution, namely the *Doppler parameter* $b_v = \sqrt{2}\sigma_v$, so that:

$$p_v = \frac{1}{\sqrt{\pi} b_v} \exp\left(\frac{-(v - v_0)^2}{b_v^2}\right). \tag{1.7}$$

In essence, eqn. 1.7 describes the probability of encountering an atom with velocity v.



In the case of a typical Lyα absorber in the IGM, the broadening of the thermal profile b_v should be of the order of $13\,\mathrm{km\,s^{-1}}$ for hydrogen atoms at $10\,000\,\mathrm{K}$.

The Intrinsic Line Profile

One of the fundamental principles of quantum mechanics, the time-energy uncertainty relation, relates the lifetime of an excited state to the uncertainty of the state's energy. More specifically, an excited state with a long lifetime will have a small energy uncertainty which results in a narrow emission or absorption profile.

Quantum mechanic uncertainty in the energy levels of transitions allow for a range of photon energies to excite an electron from a lower state l to an upper state u (with a certain probability). This cross-section for the electron-photon scattering is described as:

$$\sigma_\nu = \left(\frac{\pi e^2}{m_e c}\right) f_{lu} \phi^{\mathrm{intr}}, \tag{1.8}$$

where c is the speed of light, m_e is the mass of an electron and e is the elementary charge. The amplitude of the probability for the transition is given by the *oscillator strength* f_{lu}. Note that f_{lu} is characteristic for each transition and its value for the Lyα transition was measured in the laboratory $f_{lu,\mathrm{ly}\alpha} = 0.4146$ (Draine 2011; Meiksin 2009). The line profile ϕ^{intr} is a function that describes the probability for a photon with a certain frequency (energy) to excite the electron.

It can be shown that ϕ^{intr} can be generalized to a Lorentzian profile for atoms at rest. Its form is:

$$\phi^{\mathrm{intr}} = \frac{4\gamma_{ul}}{16\pi^2(\nu - \nu_{lu})^2 + \gamma_{ul}^2}. \tag{1.9}$$

This distribution is centered around the transition frequency ν_{lu} and has a scale-parameter γ_{ul}, that is dependent on the lifetime of the excited state.

In the following, we motivate that the intrinsic line profile is very narrow for the Lyα transition. The width of the intrinsic profile is often given as the full-width at half maximum:

$$\Delta\nu_{\mathrm{FWHM}}^{\mathrm{intr}} = \frac{\gamma_{ul}}{2\pi}. \tag{1.10}$$

One can easily transform the frequency ν_{ul} into a velocity scale by using $\Delta v_{\mathrm{FWHM}}^{\mathrm{intr}}/c = \Delta\nu_{\mathrm{FWHM}}^{\mathrm{intr}}/\nu_{ul}$.

$$\Delta v_{\mathrm{FWHM}}^{\mathrm{intr}} = \frac{c}{\nu_{ul}} \Delta\nu_{\mathrm{FWHM}}^{\mathrm{intr}} = \frac{\lambda_{ul}\gamma_{ul}}{2\pi} \tag{1.11}$$

If we use laboratory values of $\lambda_{ul}\gamma_{ul} = \lambda_{\mathrm{ly}\alpha}\gamma_{\mathrm{ly}\alpha} \simeq 7616\mathrm{cm/s}$ for the Lyα transition (Draine 2011), this will result in $\Delta v_{\mathrm{FWHM}}^{\mathrm{intr}} = 0.0121\,\mathrm{km\,s^{-1}}$. The intrinsic line width is clearly smaller than the typical thermal broadening of Lyα absorption profiles in the IGM described above. The implications of this fact will be discussed in the next section.

1.4.2. The Voigt-Profile

The Voigt profile described in eqn. 1.3, can be expressed by combining eqns. 1.6 (thermal line profile) and 1.9 (intrinsic line profile):

$$
\begin{aligned}
\phi_\nu^{\mathrm{Voigt}} &= \int p_v \cdot \phi^{\mathrm{intr}} \, \mathrm{d}v \\
&= \frac{1}{\sqrt{2\pi}} \int \frac{\mathrm{d}v}{\sigma_v} \exp\left(\frac{-v^2}{2\sigma_v^2}\right) \frac{4\gamma_{ul}}{16\pi^2(\nu - (1 - v/c)\nu_{ul})^2 + \gamma_{ul}^2} .
\end{aligned}
\tag{1.12}
$$

To a good approximation, one can treat the intrinsic profile with a δ-function, as the scale parameter of the intrinsic line profile is typically much smaller than the thermal broadening of Lyα absorbers in the IGM (see § 1.4.1). The resulting line profile can be a approximated as

$$
\phi_\nu \approx \frac{1}{\sqrt{\pi}} \frac{1}{\nu_{ul}} \frac{c}{b_v} \exp\left(-\frac{v^2}{b_v^2}\right) , \tag{1.13}
$$

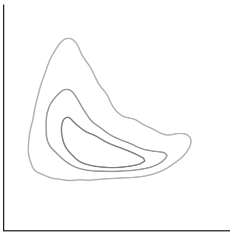
which is again Gaussian shaped. This is illustrated in the lower panel of Figure 1.5. A Gaussian approximation is appropriate for most of the optically thin ($\tau \lesssim 3$) Lyα absorbers in the IGM, but fails for high column-density absorbers, since treating the intrinsic profile as a δ-function does not account for the prominent wings of the Lorentzian profile (this is discussed in more detail in § 1.4.4).

1.4.3. Transmission in the IGM

Knowing the shape of the scattering cross-section, we can now explore how one can parameterize the shape of an absorption profile in an observed transmission spectrum.

Optical Depth

The previous section showed that the scattering probability as a function of frequency follows a very specific profile (eqn. 1.12). Observationally, it is interesting to determine how the absorption profile would appear in a spectrum in the context of the emission of a background continuum source (a QSO) passing through a medium (neutral hydrogen in the IGM). The quantity that describes the opacity of an intervening medium is the so-called optical depth τ. Given the frequency dependency of the scattering process' cross-section (see eqn. 1.8), one also expects the resulting optical depth to be a function of frequency. Furthermore, the opacity of the gas will depend on the column-density of gas in the lower state N_l, in other words, the more atoms in the line of sight, the more likely it is for absorption to take place. With this in mind and under the assumption



that stimulated emission can be neglected in the IGM, the optical depth as a function of frequency can be written as:

$$\tau_\nu = \sigma_{lu}(\nu)N_l = \frac{\pi e^2}{m_e c} f_{lu}\phi_\nu N_l. \tag{1.14}$$

Using the approximation for the line profile from eqn. 1.13, and $\nu_{ul} = c/\lambda_{ul}$, this equation can be rearranged to

$$\tau_\nu = \tau_0 \exp\left(-\frac{v^2}{b^2}\right). \tag{1.15}$$

The velocity v corresponds to the frequency shift $v = c \cdot (\nu_{lu} - \nu)/\nu_{lu}$. The optical depth at line center τ_0, is defined as:

$$\tau_0 = \frac{\sqrt{\pi} e^2}{m_e c} f_{lu}\lambda_{lu}\frac{N_l}{b}. \tag{1.16}$$

For the Lyα transition in particular, using the values $\lambda_{lu} = \lambda_{\mathrm{ly}\alpha} = 1215.67\text{Å}$ and oscillator strength $f_{lu} = 0.4164$, this equation can be written as

$$\tau_{0,\mathrm{Ly}\alpha} \simeq 0.7580 \left(\frac{N_{\mathrm{HI}}}{10^{13}\mathrm{cm}^{-2}}\right)\left(\frac{10\,\mathrm{km\ s}^{-1}}{b}\right). \tag{1.17}$$

This equation already illustrates some of the most important effects of temperature and density on the optical depth at the profile center. A high column-density N_{HI} will result in high τ_0 and increasing the Doppler parameter b will decrease τ_0.

Transmission

Once the optical depth profile for a scattering process is known, it can be easily transformed into a transmission spectrum T_ν. The observed transmission is simply

$$T_\nu = \frac{F_\nu}{C_\nu} = \exp\left(-\tau_\nu\right), \tag{1.18}$$

where C_ν is the original non-absorbed flux spectrum (the continuum) and F_ν is the observed flux spectrum. It is also common to refer to T_ν as the continuum normalized spectrum.

Note that so far the central position of the line has been treated as being zero in a velocity scale. A shift in line-of-sight velocity v_0 can be added by simply using $v - v_0$ instead of v in the equations above. In terms of redshift, this corresponds to an absorber with observed wavelength $\lambda_0 = \lambda_{\mathrm{Ly}\alpha} \cdot (1 + z)$ and it follows that $v - v_0 = (\lambda - \lambda_0)/\lambda_0$. So the position of the line in a given spectrum can be described in terms of redshift.

Altogether, an absorption profile in a transmission spectrum can be parameterized by a column-density N_l of the absorbing gas, a thermal[7] broadening Doppler parameter b and its position z in redshift space.

1.4.4. The Curve of Growth

In this section we will explore how the column density of an absorber affects the absorption profile in a spectrum. To this end, I will use one measure for the absorbed flux in a spectrum which is given by the equivalent width:

$$W = \int \frac{\mathrm{d}\nu}{\nu_0}\left(1 - \frac{F_\nu}{C_\nu}\right) = \int \frac{\mathrm{d}\nu}{\nu_0}(1 - \exp(-\tau_\nu)), \tag{1.19}$$

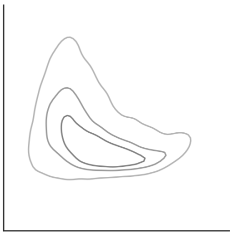
where ν_0 is the central frequency of the absorption profile. The equivalent width is defined so that a rectangular integral of the flux over the width W corresponds to the integral over the line profile. The dependency of the equivalent width on the column density N is described by the so-called curve of growth.

The curve of growth is divided in three main regimes, which are separated by the different sensitivities of W on N. The three main cases are shown in Figure 1.6.

- When the absorption profile has a relatively small τ_0 (black line), then $W \propto N$, i.e. the equivalent width will increase linearly with column density. This linear regime corresponds to most of the absorption caused by optically thin lines in the IGM with $12 \lesssim \log(N_{\mathrm{HI}}/\mathrm{cm}^{-2}) \lesssim 14$.

- Once the optical depth at line center increases and absorption features start to saturate (gray line), the curve of growth enters the flat regime. Typically this is the case for lines in the IGM with $\log(N_{\mathrm{HI}}/\mathrm{cm}^{-2}) \gtrsim 14$ (with the exact point of turnover being dependent on b). Due to the nearly rectangular shape of the absorption, increasing the column density does not affect the line profile significantly, resulting in a dependency $W \propto \sqrt{\ln N}$. Interestingly, this causes the shape of lines to become strongly sensitive to changes in b.

- For higher values of N, the Lorentzian wings of the Voigt profile become increasingly important (light gray line). Although the scattering cross section is nearly Gaussian, the exponentiation of the optical depth used to generate the flux (see eqn. 1.18) will amplify these subtle differences. Damped absorbers in the IGM typically have column densities of $\log(N_{\mathrm{HI}}/\mathrm{cm}^{-2}) \gtrsim 20$. Due to the Lorentzian wings, the column density is discernible and becomes sensitive to the equivalent width as $W \propto \sqrt{N}$.



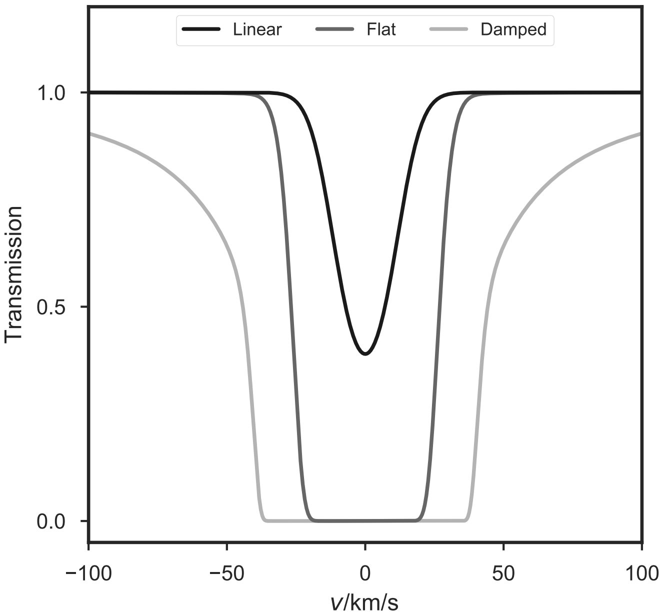


Figure 1.6. Illustration of the three cases of the curve of growth. The black line represents an absorption line with $\tau_0 \simeq 1$ as an example for the linear part of the curve of growth. An example of a saturated line with $\tau_0 \simeq 50$ in the flat regime is shown in dark gray. Damped absorption with $\tau_0 \simeq 10^4$ is shown in light gray. The Voigt profile used for generating these absorption features is the same as shown in the lower panel of Figure 1.5. The absorption profile was transformed into flux using eqn. 1.18.

The curve of growth sets a limitation on how well one is able to measure absorption line parameters from a particular line transition. However, if one combines different transitions of the same gas, that have different oscillator strengths and therefore different column density thresholds for the different dependencies of the curve of growth, one can still determine line parameters. In this study I will focus on the Lyα transition alone, as I will rely on automatization of the line fitting procedure in the upcoming chapters, which becomes substantially more complicated once higher transitions are taken into account. I will also concentrate on absorbers with $11.5 \lesssim \log(N_{\mathrm{HI}}/\mathrm{cm}^{-2}) \lesssim 16$, which correspond to the diffuse low density IGM.

In the literature one separates between IGM absorbers, Lyman limit systems (LLS) and damped Lyα (DLA) systems. LLSs[8] are absorbers with $\log(N_{\mathrm{HI}}/\mathrm{cm}^{-2}) \gtrsim 17$, that are in the flat regime of the curve of growth. DLAs are absorbers with $\log(N_{\mathrm{HI}}/\mathrm{cm}^{-2}) \gtrsim 20.3$

---

[7] We will discuss in § 1.5 that there are other contributions to the broadening of a line, but that these can be treated jointly with one single Doppler parameter.

[8] The name Lyman limit system refers to the lack of transmission after the Lyman limit that is associated with these absorbers due to self-shielding. See discussion in the text.

that show prominent Lorentzian wings and are in the damped regime. These high column density absorbers are self-shielding, i.e. they have high enough enough density to preserve a neutral hydrogen core that is protected from the UVB. In the case of DLAs these hydrogen column densities are comparable to that of the Milky Way (MW), making them presumably star forming (Fumagalli et al. 2014) proto-galaxies in the line of sight.

These high density objects have to be carefully removed from our sample, as we are interested in measuring the thermal state of the low density IGM. Damped systems are easily identified due to the Lorentzian wings and associated metal line absorption. LLS, in contrast, are not easily discriminated from strong absorbers or blended absorbers originating in the IGM. Most of time, LLS are detected via corresponding metal absorption. If the spectral coverage is large enough, self-shielding absorbers will completely absorb photons blueward of the Lyman limit (rest frame 912 Å, shifted to the absorber's position in redshift), due to the high scattering probability at such high densities. Therefore, these systems can also be identified by complete absorption at short wavelengths.

In fact, this is visible in Figure 1.4 at 3410 Å, which indicates that there is a self shielding absorber at $z = 2.74$ causing Lyman limit absorption. The corresponding Lyα absorber should be at 4546 Å. When inspecting this part of the spectrum in Figure 1.7, we find the Lyα signature of these absorbers.

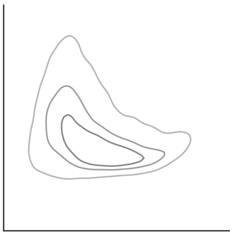
I will discuss in detail how we remove these objects from observed datasets in § 2.1.4.

1.5. The b-$N_{\mathrm{H\,I}}$ distribution

In this thesis I will explore different methods of extracting the parameters that govern the TDR in the IGM from the distribution of absorption line parameters of Lyα absorbers in the Lyα forest (see § 1.3). As described in the previous sections, absorbers in the IGM can be described by their position in redshift space z_{abs}, their column-density $N_{\mathrm{H\,I}}$ and their Doppler parameter b. I will explore in detail how the joint distribution of $N_{\mathrm{H\,I}}$ and b (the b-$N_{\mathrm{H\,I}}$ distribution) is sensitive to the thermal state of the IGM and how it evolves as a function of time (redshift).

1.5.1. Thermal Sensitivity of the b-$N_{\mathrm{H\,I}}$ distribution

The b-$N_{\mathrm{H\,I}}$ distribution is one of the primary observables that are sensitive to the thermal state of the IGM, because the column-density $N_{\mathrm{H\,I}}$ can be treated as a proxy for the neutral hydrogen density and the Doppler parameter can be treated as a proxy for temperature. In detail, the Doppler parameter b of an absorber is determined by the contributions from its thermal state, but is expected to have additional contributions.



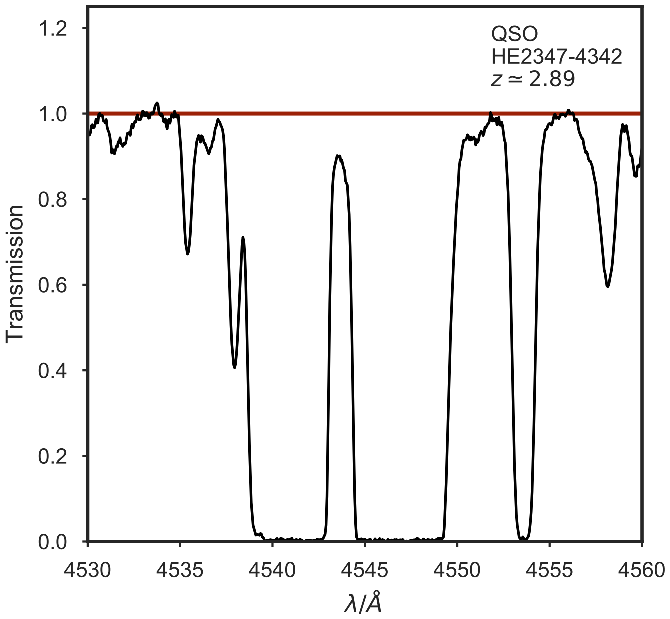


Figure 1.7. The same spectrum as shown in Figure 1.4, normalized by its continuum, and showing only the region around 4546 Å, where total absorption beyond the Lyman limit indicates that there should be a self-shielding Lyα absorber. In fact, we see two neighboring LLS, the rightmost of which causes the lack of transmission, leaving no Lyman-limit photons to be absorbed by the rightmost. These absorbers have no prominent Lorentzian wings and are therefore not DLAs.

As discussed earlier, the thermal contribution consists of microscopic random thermal motions in the gas, or *thermal broadening*

$$b_T = \sqrt{\frac{2k_BT}{m_H}} \tag{1.20}$$

and is simply a result of blue and redshifting of the absorption wavelength due to Maxwell-Boltzmann velocity distributions in the gas. In addition to thermal broadening, one also expects the kinematic structure in the absorber to affect the width of absorption profiles. This is often (rather confusingly) referred to as *turbulent broadening* $b_{\mathrm{turb.}}$.

Additionally, the differential Hubble flow across the spatial extent of an absorbing cloud, which is set by the so-called pressure smoothing scale λ_P (Gnedin & Hui 1998; Schaye 2001; Rorai et al. 2013; Kulkarni et al. 2015; Rorai et al. 2017b) will contribute to the line widths. This effect is often referred to as Hubble broadening b_H. One can motivate λ_P by assuming that the overdensities probed by the Lyα forest are close to hydrostatic equilibrium (Schaye 2001). This means that the sound crossing time, which is the timescale of the pressure support $t_{\rm sc} = L/c_{\rm s}$, is comparable to the free-fall time $t_{\rm ff} = 1/\sqrt{G\rho}$, with L being the radial extent of the cloud, c_s the sound speed and G is the gravitational constant. This condition results in a characteristic smoothing scale, the Jeans-scale $\lambda_J \coloneqq L = c_{\rm s}/\sqrt{G\rho}$, which is related to λ_P (see definition of λ_P by Kulkarni et al. 2015, discussed in § 2.2). Hubble broadening due to the spatial extent of an absorber scales as $b_H \sim H(z)\,\lambda_J$, where $H(z)$ is the Hubble parameter. This broadening mechanism is expected to be comparable to the thermal broadening at mean density and should be taken into account Garzilli et al. (2015, 2018).

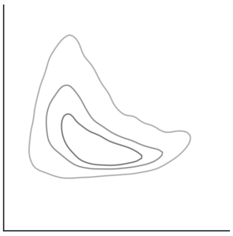
To a good approximation, the total Doppler parameter of an absorber is given by

$$b^2_{\rm tot.} = b^2_T + b^2_{\rm turb.} + b^2_H. \tag{1.21}$$

Figure 1.8 illustrates the thermal sensitivity of the b-$N_{\rm H\,I}$ distribution, based on the output of Voigt profile fitting of simulated sightlines, i.e. with known thermal parameters. Note that instead of showing a cloud of points we plot a color coded map with the probability $P(\log N_{\rm H\,I}, \log b)$ (labeled as PDF) for better visibility. The method used to generate these maps will be discussed in detail in Chapter 3.

The left and center panels of Figure 1.8 show typical b-$N_{\rm H\,I}$ distributions of the intergalactic gas. This distribution has many interesting features that will be discussed in this thesis. For one, we observe that there is a sharp *cutoff* towards low b values in all distributions. This cutoff comes from the lower b limit at each density that is imposed by the temperature of the gas. Because of all the other contributions to the broadening of absorbers, we see a spread in b toward higher values. However, the cutoff simply reflects that some absorbers might have negligible kinematic contribution if the peculiar velocities have no contribution to the broadening along the line of sight. At very low $N_{\rm H\,I}$ the cutoff has a turnaround toward higher b. As discussed in Garzilli et al. (2015, 2018) this is caused by Hubble broadening across the spatial extent of the absorber becoming comparable to the thermal broadening for the lowest column density absorbers. One also observes a diagonal cutoff towards low $N_{\rm H\,I}$. This is simply an observational effect. As discussed earlier, the optical depth at line-center is $\tau_{0,\rm H\,I} \propto N_{\rm H\,I}/b$ which illustrates that lines that have low $N_{\rm H\,I}$ and/or high b will be less likely to be detected depending on the SNR of the spectra at hand.

The top panels of Figure 1.8 show how the b-$N_{\rm H\,I}$ distribution reacts to a change of T_0 in our simulations. One sees that increasing the temperature at mean-density shifts the cutoff of the distribution toward higher b values. The bottom panels show the effect of changing γ. In the TDR, γ regulates the contrast in temperature between over and underdensities. This is also reflected in the b-$N_{\rm H\,I}$ distribution. Increasing γ results in a tilt of the lower cutoff of the distribution, as overdensities (higher $N_{\rm H\,I}$) become hotter.



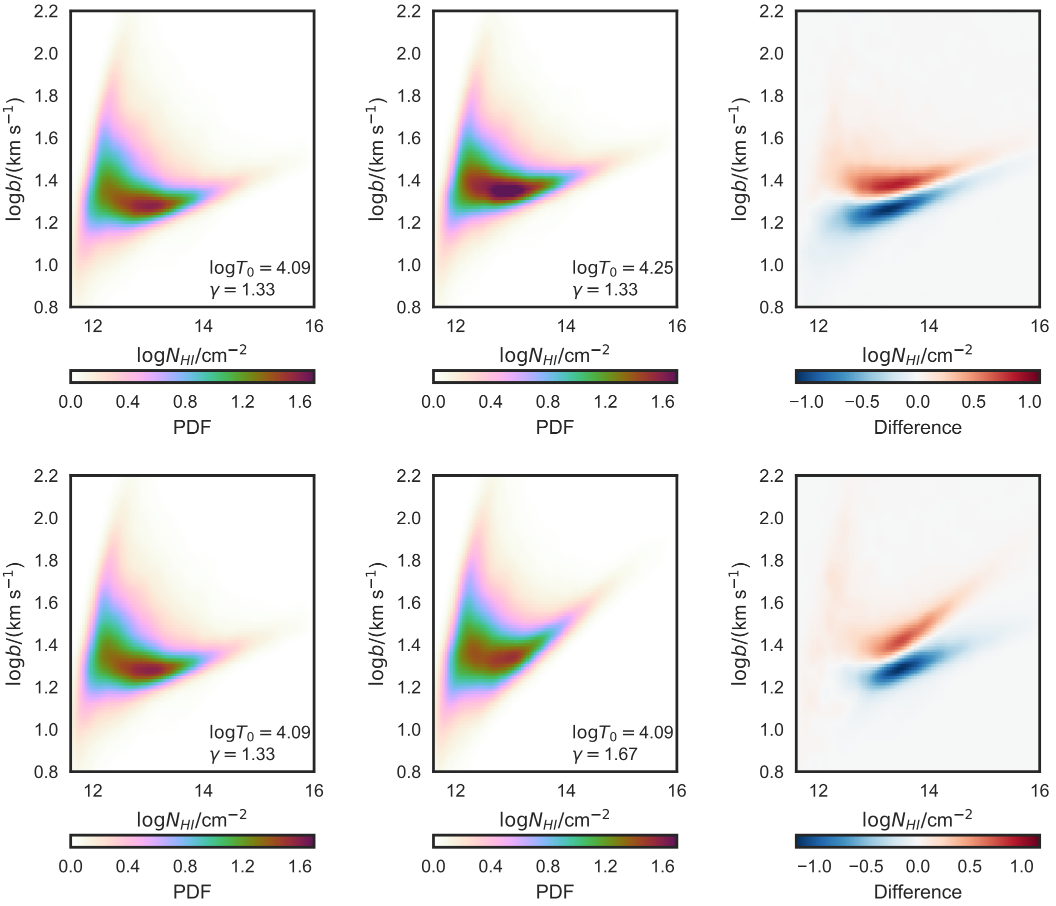


Figure 1.8. Sensitivity of the b-N_{HI} distribution on the TDR parameters imposed onto simulations. Note that instead of showing a cloud of points we plot a color coded map with the probability $P(\log N_{\mathrm{HI}}, \log b)$ (labeled as PDF) for better visibility. **Left:** $P(\log N_{\mathrm{HI}}, \log b)$ from as simulation with $\log T_0 = 4.09$ and $\gamma = 1.33$. **Middle:** $P(\log N_{\mathrm{HI}}, \log b)$ from the same simulation but with higher $\log T_0$ (bottom) and higher γ (bottom). **Right:** The difference between the distributions illustrates that increasing $\log T_0$ shifts $P(\log N_{\mathrm{HI}}, \log b)$ toward higher b (top), while increasing γ mainly tilts the distribution at $\log(N_{\mathrm{HI}}/\mathrm{cm}^{-2})$> 13 (bottom).

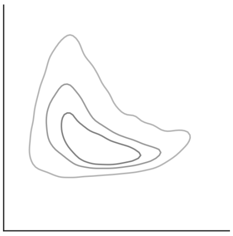
1.6. Scope of This Thesis

In this thesis I will carry out measurements of T_0 and γ as a function of redshift. This will be done by carefully comparing the b-$N_{\mathrm{H\,I}}$ distributions retrieved from observational data to the b and $N_{\mathrm{H\,I}}$ output of state of the art hydrodynamic simulations. All detailed statistical methods, data, simulations, manipulations and results will be discussed in three different chapters:

- **In Chapter 2** I present a new version of a classic approach of measuring thermal parameters from the b-$N_{\mathrm{H\,I}}$ distribution, namely by quantifying the position of the lower cutoff in the distribution. I expand on previous work by significantly increasing the number of observed sightlines used (75 Ultraviolet and Visual Echelle Spectrograph (UVES) and High Resolution Echelle Spectrometer (HIRES) spectra) and carefully calibrating the relationship between cutoff and thermal parameters with the help of hydrodynamic simulations. Furthermore, consideration is given to the effect of the pressure smoothing scale λ_P on the uncertainty of our measurements. With this method I measure the evolution of T_0 and γ in the redshift range $2 \leq z \leq 3.4$. The results of this work have been published in Hiss et al. (2018).

- **In Chapter 3** I develop, test and apply a new method of carrying out measurements of thermal parameters using the b-$N_{\mathrm{H\,I}}$ distribution that goes beyond the cutoff and instead treats the b-$N_{\mathrm{H\,I}}$ distribution as a whole. This approach is more accurate when compared with cutoff fitting and delivers measurements with significantly smaller statistical uncertainties for current dataset sizes. This method is applied to observational data at $z = 2$. The results of this work are published in Hiss et al. (2019).

- **In Chapter 4** I measure the thermal state of the IGM at $z = 0.1$ by applying the method of Chapter 3 to spectra from the Hubble Space Telescope (HST)/Cosmic Origins Spectrograph (COS) instrument. This study is in preparation and its results will be submitted for publication in the near future.



2. Fitting the Cutoff of the b-$N_{\mathrm{H\,I}}$ Distribution at $2 \leq z \leq 3.4$

The color of the sun composes me
The blue sea dissolves me
The equation proposes me
The computer solves me

- Os Mutantes, *2001* 1969

In this Chapter we will apply the cutoff fitting method discussed in § 1.5.1 in order to carry out a measurement of the parameters T_0 and γ of the TDR. This method is based on the idea that, if we observe many absorption features, we will occasionally encounter lines from gas clouds which have a line of sight velocity component near zero, i.e. the velocity field is close to a caustic (McDonald et al. 2001). As the broadening of these absorbers is dominated by the thermal contribution, this results in a thermal state dependent cutoff in the distribution of Doppler parameters of the form

$$b_{\mathrm{min.}} = b_0 \left(\frac{N_{\mathrm{H\,I}}}{N_{\mathrm{H\,I},0}} \right)^{\Gamma - 1}, \tag{2.1}$$

where b_0 is the minimal Doppler parameter at the column density $N_{\mathrm{H\,I},0}$ and Γ is the power law index that regulates the contrast in minimal broadening for different column densities. The quantity $N_{\mathrm{H\,I},0}$ refers to the column density that corresponds to an absorber with the mean density ρ_0. The formalism behind this cutoff form will be discussed in more detail in § 2.3.1.

Assuming that the cutoff is primarily set by the thermal state of the gas, its position will be dependent on the gas density due to the temperature-density relation, or in observable terms, the absorption-line column density $N_{\mathrm{H\,I}}$. This in turn means that there is a correlation between the position of the lower cutoff in the distribution of Doppler parameters as a function of column densities (b-$N_{\mathrm{H\,I}}$ distribution) and the thermal state of the gas. The strategy we will adopt is to calibrate the relationship between b_0 and T_0, as well as Γ and γ using the Voigt profile fit output of hydrodynamic simulations. Once the calibration is known, we can apply it to cutoff measurements in the b-$N_{\mathrm{H\,I}}$ distributions of observational data, allowing us to constrain T_0 and γ.

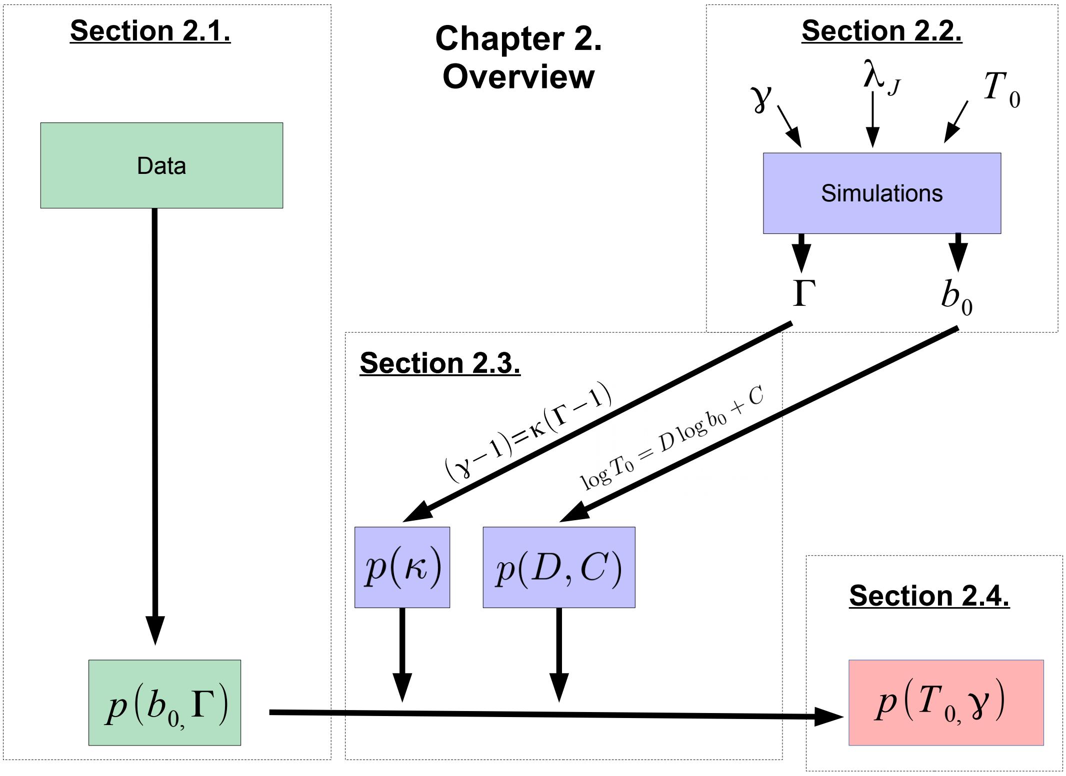


Figure 2.1. Overview of this chapter.

This method has been explored in the literature (see Schaye et al. 1999; Ricotti et al. 2000; McDonald et al. 2001), but recent development in both the observational and computational front pointed for the need of revisiting cutoff fitting techniques. A few years ago, a measurement of the thermal state of the IGM based on the cutoff of the b-N_{HI} distribution was performed at $z = 2.4$ by Rudie et al. (2012a) using a sample of 15 high-quality QSO sightlines. Using the analytic relations between the cutoff of the b-N_{HI} distribution and the temperature-density relation derived by Schaye et al. (1999), Rudie et al. (2012a) measured a temperature at mean density $T_0 = 1.94 \times 10^4\,\mathrm{K}$ at $z \sim 2.4$ (in a broad redshift bin spanning $2.0 < z < 2.8$), which was $\sim 9000\,\mathrm{K}$ higher than the value implied by flux curvature measurements at $z = 2.4$ by Becker et al. (2011). This discrepancy motivated Bolton et al. (2014) to revisit the cutoff fitting technique. Using hydrodynamic simulations, they calibrated the relationship between the b-N_{HI} distribution cutoff and the temperature-density relation. Applying this updated calibration to the Rudie et al. (2012a) b-N_{HI} distribution cutoff measurement, Bolton et al. (2014) determined a lower temperature $T_0(z = 2.4) = [1.00^{+0.32}_{-0.21}] \times 10^4\,\mathrm{K}$ that is consistent with Becker et al. (2011), and argued that the much higher temperature

measured by Rudie et al. (2012a) resulted from incorrect assumptions in the calibration.

With the knowledge that calibration using simulations yields more reliable measurements, we carried out this study of the b-$N_{\mathrm{H\,I}}$ distribution of an extensive sample of 75 high quality QSO spectra, which allows us to measure the redshift evolution of T_0 and γ over the redshift range $2.0 \leq z \leq 3.4$ with a much finer binning $\delta z = 0.2$ than previous work. At each redshift we use mock Lyα forest data from 26 hydrodynamic simulations with different thermal histories to calibrate the relationship between the cutoff in the b-$N_{\mathrm{H\,I}}$ distribution and the thermal parameters (T_0, γ, λ_P) governing the IGM. The Lyα forest of both the data and the simulations are decomposed into individual absorption lines using the Voigt profile fitting algorithm `VPFIT` (Carswell & Webb 2014), and we adopt a forward-modeling approach whereby the same algorithms are self-consistently applied to both data and simulations.

This chapter is structured as follows (see Figure 2.1). We introduce our dataset, Voigt-profile, and cutoff fitting procedure and resulting probability density function (PDF)s $p\,(b_0, \Gamma)$ in § 2.1. An overview of our hydrodynamic simulations, with known parameters T_0, γ and pressure scale λ_P, as well how we measure b_0 and γ from them is described in § 2.2. In this section we also introduce the THERMAL (Thermal History and Evolution in Reionization Models of Absorption Lines) suite. In § 2.3 we discuss how we calibrate our method, i.e. how we relate b_0 to T_0 and Γ to γ, by applying the same fitting procedures to simulated sightlines. Our final result, a Monte Carlo combination of the calibration values $p(\kappa)$, $p(D, C)$ and the cutoff measurements $p\,(b_0, \Gamma)$, is a measurement of the evolution of the thermal state $p(T_0, \gamma)$ of the IGM at $2 \leq z \leq 3.4$. It is presented and discussed in § 2.4. We summarize our results in § 2.5.

2.1. Data Processing

2.1.1. QSO Sample

For this study, we used a sample of 75 publicly available QSO spectra with SNR better than 20 per 6 $\mathrm{km\,s^{-1}}$ bin and resolution varying between $\mathrm{FWHM} = 3.1$ and 6.3 $\mathrm{km\,s^{-1}}$ with a typical value around 6 $\mathrm{km\,s^{-1}}$. This ensures that the Lyα forest is resolved and that we can detect lines with $N_{\mathrm{H\,I}} \simeq 10^{12.5}$ $\mathrm{cm^{-2}}$ at the 3σ level (Herbert-Fort et al. 2006). Part of the sample consists of QSO spectra from the Keck Observatory Database of Ionized Absorbers toward QSOs (KODIAQ) sample (Lehner et al. 2014; O'Meara et al. 2016, 2017). The other spectra were acquired with the UVES instrument (Dekker et al. 2000) at the Very Large Telescope (Dall'Aglio et al. 2008).

The KODIAQ sample used in this work consists of 36 QSO sightlines chosen from DR1 and DR2. These QSOs were observed between 1995 and 2012 using the HIRES (High Resolution Echelle Spectrometer: Vogt et al. 1994) instrument on the Keck-I telescope. All the spectra were homogeneously reduced and continuum-fitted by the KODIAQ team using the HIRedux code[1]. Multiple exposures of the same spectra were co-added in order to increase the SNR. The detailed reduction steps are described in O'Meara et al. (2015).

The UVES spectra consist of 38 sightlines from the ESO archive. These objects were chosen to have complete, or nearly complete, Lyα forest coverage and at least 10 exposures each. The data were reduced by Dall'Aglio et al. (2008) using the `MIDAS` environment ECHELLE/UVES and procedures described in Kim et al. (2004). Each frame was bias- and background-subtracted. The Echelle spectra were extracted order by order assuming a Gaussian profile along the spatial direction. The final spectra have very high SNR per pixel ≥ 40 and a resolution of 6 km s^{-1} in the Lyα forest region after co-adding. Continua were fitted by Dall'Aglio et al. (2008) using a cubic-spline interpolation method. We used 38 spectra from the 40 available in this sample. One characteristic of the UVES pipeline is that the estimated errors at flux values close to zero are underestimated by a factor of roughly two (Carswell et al. 2014). Therefore, Voigt profile fitting algorithms will struggle to achieve a satisfactory χ^2 for these regions. To compensate for this, we used a dedicated tool implemented in RDGEN[2] (Carswell et al. 2014), a front- and back-end program for `VPFIT`. This tool multiplies the error of each pixel with a value that is 1 if the corresponding normalized flux is 1 and 2 if the normalized flux is 0. For this purpose, we used the default parameterization from RDGEN.

The region of the spectra used for fitting lies between 1050 and 1180 Å rest frame inside the Lyα forest. This region was chosen to avoid proximity effects, i.e. regions dominated by the local QSO radiation rather than the metagalactic UVB. This choice is consistent with studies by Palanque-Delabrouille et al. (2013) and Walther et al. (2019).

For a complete list of the spectra analyzed in this work and the essential information about them, refer to Tables 2.1 and 2.2. The chunks of spectra used are plotted in Figure 2.2 and colored based on the dataset they belong to. Our analysis of the thermal state of the IGM will be done in redshift bins of size $\delta z = 0.2$, indicated with vertical blue lines. We discuss the effects of continuum misplacement in our data in the appendix A.1.

2.1.2. Voigt-Profile Fitting

Voigt profiles are fitted to our data using `VPFIT` version 10.2[3] (Carswell & Webb 2014). We wrote a fully automated set of wrapper routines that prepares the spectra for the fitting procedure and controls `VPFIT` with the help of the front-end/back-end programs

---

[1] HIRedux: http://www.ucolick.org/~xavier/HIRedux/

[2] RDGEN:http://www.ast.cam.ac.uk/~rfc/rdgen.html

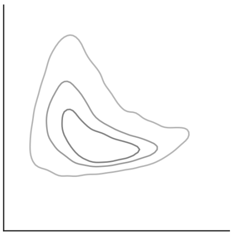
[3] VPFIT: http://www.ast.cam.ac.uk/~rfc/vpfit.html

Table 2.1. QSO spectra used in this work. The signal-to-noise value refers to the median value inside the Lyα forest.

Object ID	z_{qso}	$SNR/6\ \mathrm{kms}^{-1}$	Sample
HE1341-1020	2.137	58	UVES
Q0122-380	2.192	56	UVES
J122824+312837	2.2	87	KODIAQ
J110610+640009	2.203	59	KODIAQ
PKS1448-232	2.222	57	UVES
PKS0237-23	2.224	102	UVES
HE0001-2340	2.278	66	UVES
J162645+642655	2.32	104	KODIAQ
J141906+592312	2.321	37	KODIAQ
Q0109-3518	2.406	70	UVES
HE1122-1648	2.407	172	UVES
HE2217-2818	2.414	94	UVES
Q0329-385	2.437	58	UVES
HE1158-1843	2.459	67	UVES
J005814+011530	2.495	36	KODIAQ
J162548+264658	2.518	44	KODIAQ
J121117+042222	2.526	34	KODIAQ
J101723-204658	2.545	70	KODIAQ
Q2206-1958	2.567	75	UVES
J234628+124859	2.573	75	KODIAQ
Q1232+0815	2.575	46	UVES
HE1347-2457	2.615	62	UVES
J101155+294141	2.62	130	KODIAQ
J082107+310751	2.625	64	KODIAQ
HS1140+2711	2.628	89	UVES
J121930+494052	2.633	90	KODIAQ
J143500+535953	2.635	65	KODIAQ
Q0453-423	2.663	78	UVES
J144453+291905	2.669	134	KODIAQ
PKS0329-255	2.705	48	UVES
J081240+320808	2.712	49	KODIAQ
J014516-094517A	2.73	77	KODIAQ
J170100+641209	2.735	82	KODIAQ
Q1151+068	2.758	49	UVES
Q0002-422	2.768	75	UVES
HE0151-4326	2.787	98	UVES
Q0913+0715	2.788	54	UVES
J155152+191104	2.83	30	KODIAQ

Table 2.2. QSO spectra used in this work (continued).

Object ID	z_{qso}	$SNR/6\ \mathrm{kms}^{-1}$	Sample
Q1409+095	2.843	25	UVES
Q0119+1432	2.87	33	KODIAQ
J012156+144820	2.87	55	KODIAQ
Q0805+046	2.877	27	KODIAQ
HE2347-4342	2.886	152	UVES
J143316+313126	2.94	54	KODIAQ
J134544+262506	2.941	35	KODIAQ
Q1223+178	2.955	33	UVES
Q0216+08	2.996	37	UVES
HE2243-6031	3.011	119	UVES
CTQ247	3.026	69	UVES
J073621+651313	3.038	26	KODIAQ
J194455+770552	3.051	30	KODIAQ
HE0940-1050	3.089	70	UVES
J120917+113830	3.105	31	KODIAQ
Q0420-388	3.12	116	UVES
CTQ460	3.141	41	UVES
J114308+113830	3.146	32	KODIAQ
J102009+104002	3.168	36	KODIAQ
Q2139-4434	3.208	31	UVES
Q0347-3819	3.229	84	UVES
J1201+0116	3.233	30	KODIAQ
J080117+521034	3.236	43	KODIAQ
PKS2126-158	3.285	64	UVES
Q1209+0919	3.291	30	UVES
J095852+120245	3.298	45	KODIAQ
J025905+001126	3.365	26	KODIAQ
Q2355+0108	3.4	58	KODIAQ
J173352+540030	3.425	57	KODIAQ
J144516+095836	3.53	25	KODIAQ
J142438+225600	3.63	29	KODIAQ
Q0055-269	3.665	76	UVES
Q1249-0159	3.668	70	UVES
Q1621-0042	3.708	78	UVES
Q1317-0507	3.719	42	UVES
PKS2000-330	3.786	151	UVES
J193957-100241	3.787	66	KODIAQ



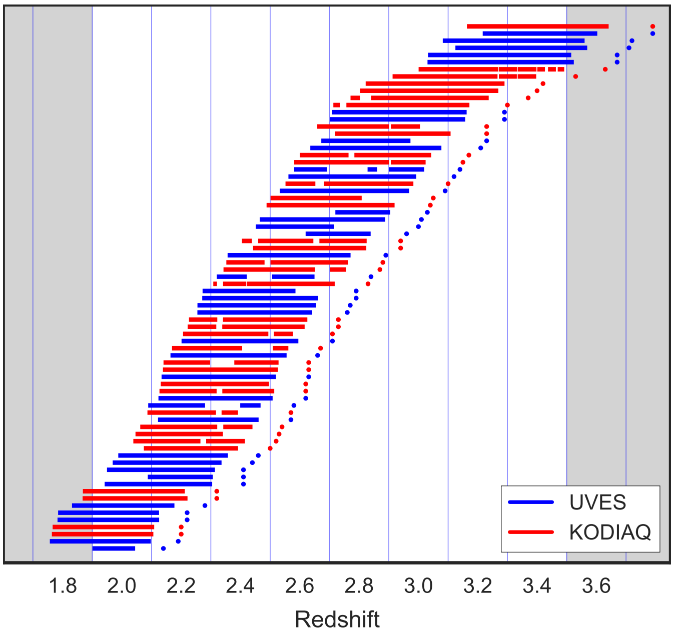


Figure 2.2. KODIAQ and UVES Lyα forest sightlines used in this work. The sample is described in § 2.1. Sightlines from the KODIAQ sample are marked in red, and UVES sightlines are marked in blue. The corresponding QSO redshifts are marked as colored points. The DLAs and bad regions that were excluded are shown as gaps. The blue vertical lines show the bins that will be used for the cutoff fitting analysis. Gray regions are not used because of lower coverage. Figure credit: Hiss et al. (2018).

RDGEN and AUTOVPIN. These are used to generate initial guesses for the absorption-line parameters, output tables and determine which segments to fit separately. For each segment VPFIT computes the best fitting superposition of absorption profiles that describes a given spectrum. Each profile is described by three parameters: line redshift $z_{\rm abs}$, Doppler parameter b, and column density $N_{\rm HI}$ corresponding to the desired absorbing gas transition (here Ly-α). The parameter space chosen for VPFIT to look for lines was set to go from 1 to 300 km s^{-1} in b and 11.5 to 16.0 in $\log(N_{\rm HI}/{\rm cm}^{-2})$. Then VPFIT varies these parameters and searches for a solution that minimizes the χ^2. It will add components until the fit converges or no longer improves. In order to minimize computational time, this fitting procedure is done in different segments of the spectra at a time. The front-end programs allow us to automatically find regions that are between sections of the spectra where the flux meets the continuum (unabsorbed regions) and fit them separately. For each spectrum VPFIT outputs a list of line parameters.

Damped Lyα systems (DLAs), i.e. Lyα absorbers with $N_{\rm HI} \gtrsim 10^{20}{\rm cm}^{-2}$, were identified by eye and are excluded from our analysis. The DLAs were chosen to enclose a

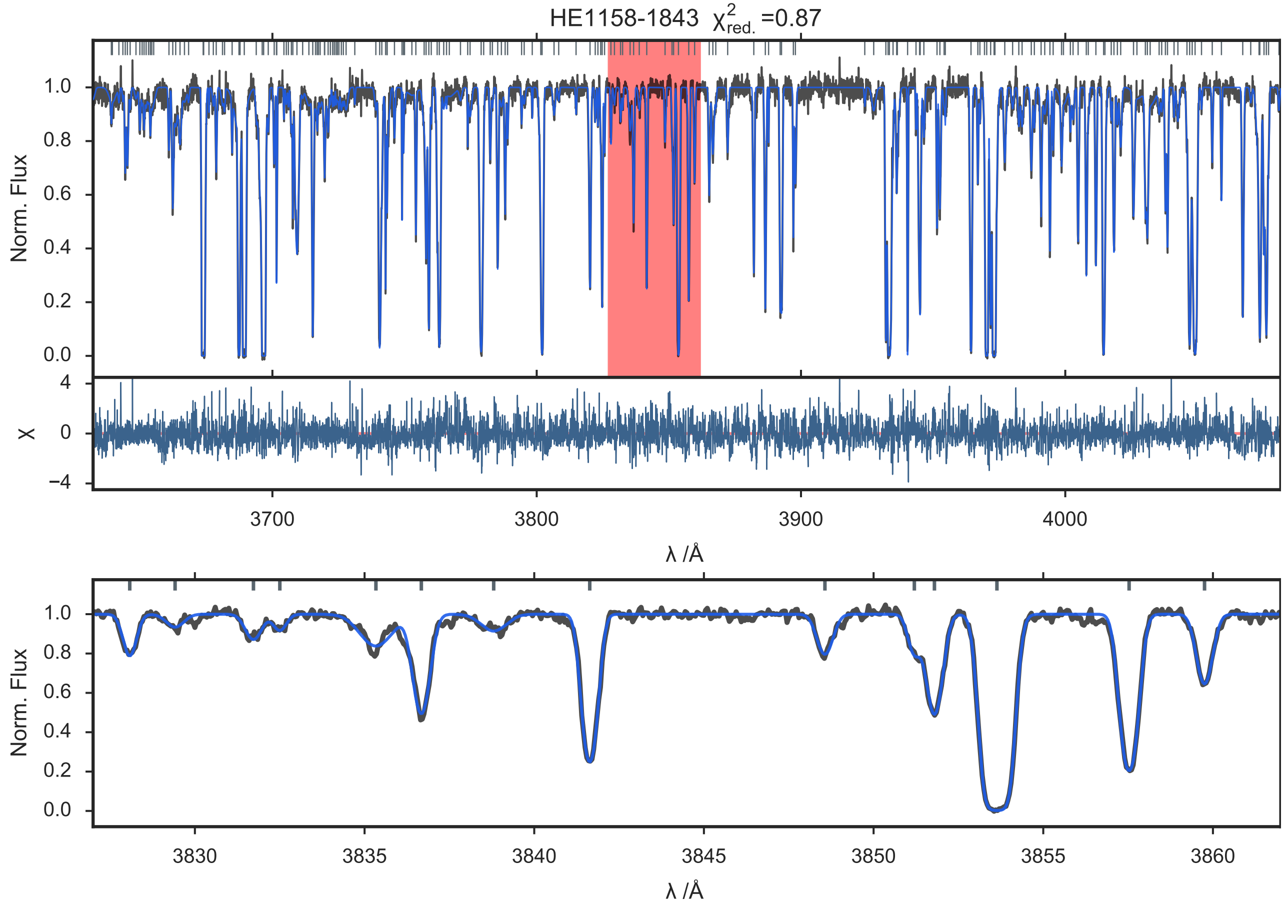
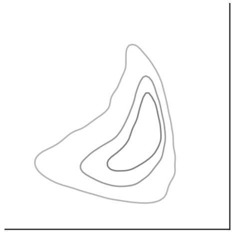


Figure 2.3. Voigt profile fitting results for the Lyα forest of the UVES QSO HE1158-1843 at $z \simeq 2.46$. **Top panel**: the measured spectrum (black line) is well described by the best fitting superposition of Voigt profiles estimated by `VPFIT` (blue line). The position of individual lines is shown by gray rugs in the top part of the panel. In the bottom panel, we plot the corresponding $\chi = (F_{\rm spec} - F_{\rm fit})/\sigma_{F_{\rm fit}}$ as a measure for the performance of the fit. **Bottom Panel:** zoom in of the area of the spectrum marked in red in the top panel.



region between the two points where the damping wings reach the QSO continuum within the flux error. Additionally, regions larger than 30 pixels previously masked in the data (bad pixels, gaps, etc.) were also excluded. We simply cut out the regions in which these rejections apply and feed the usable data segments into `VPFIT` separately.

In order to avoid chopping our spectra into too many small segments, small regions ($\leq$ 30 pixels) that were previously masked in the data were cubically interpolated. These pixels were given a flux error of a 100 times the continuum, so the Voigt profile fitting procedure is not influenced.

One complication is that `VPFIT` often has difficulty fitting the boundaries of spectra. To solve this problem, we artificially enlarge the spectral chunks. For this purpose, we append a mirrored version of the first quarter of the spectra to the beginning of it. We do the same with the last quarter to the end of the spectrum. These regions and the absorption profiles within them are later ignored. This method ensures that the unreliable fits at the boundaries happen in an artificial environment that is discarded. The disadvantage is that the spectrum that is Voigt profile fitted is 50% longer than the original and will therefore need more time to be processed.

An example of the Voigt profile fitted spectrum of a UVES sightline is shown in Figure 2.3.

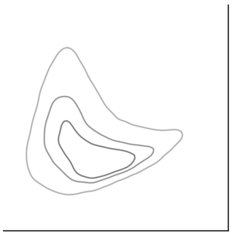
2.1.3. The b-$N_{\mathrm{H\,I}}$ Distribution

The output of `VPFIT` can be used to generate a $\log b$ vs. $\log(N_{\mathrm{H\,I}}/\mathrm{cm}^{-2})$ diagram (b-$N_{\mathrm{H\,I}}$ distribution). Note that for a comparatively small number of lines, `VPFIT` might output error estimates that are zero, `nan` or "*******". When generating diagrams, we exclude these lines, because they normally appear in blended regions and noisy parts of the spectra.

In order to illustrate the effect of SNR on the b-$N_{\mathrm{H\,I}}$ distribution, we generate 2 b-$N_{\mathrm{H\,I}}$ distributions by Voigt profile fitting mock Lyα forest absorption spectra at $z = 2.4$ with different SNR applied to them. For this simple exercise, we used mock Lyα forest spectra based on collisionless DM only simulations[4]. The resulting distributions are shown in Figure 2.4. In one case (red) we added a constant and extremely high SNR/6 kms^{-1} of 280, while in the other case (black) a SNR based on the data at $z = 2.4$ (with a median of SNR/6 kms^{-1}=64) was applied. Some of the features are identical, especially the existence and position of a cutoff at $\log N_{\mathrm{H\,I}}/\mathrm{cm}^{-2} > 12.5$ and $\log b/(\mathrm{kms}^{-1}) \sim 1.2$. The main difference is that the high SNR distribution is more complete toward low $\log N_{\mathrm{H\,I}}$ and high $\log b$ values. At column densities $\log N_{\mathrm{H\,I}}/\mathrm{cm}^{-2} > 12.5$ and Doppler

---

[4] These simulations use an updated version of the TreePM code from White et al. (2002), similar to Rorai et al. (2013, 2017b), that evolves $N_p = 2048^3$ collisionless, equal-mass particles ($M_p = 2.5 \times 10^5 \mathrm{M}_\odot$) in a periodic cube of side length $L_{\mathrm{box}} = 30\,\mathrm{Mpc}/h$, adopting a Planck Collaboration et al. (2014) cosmology.



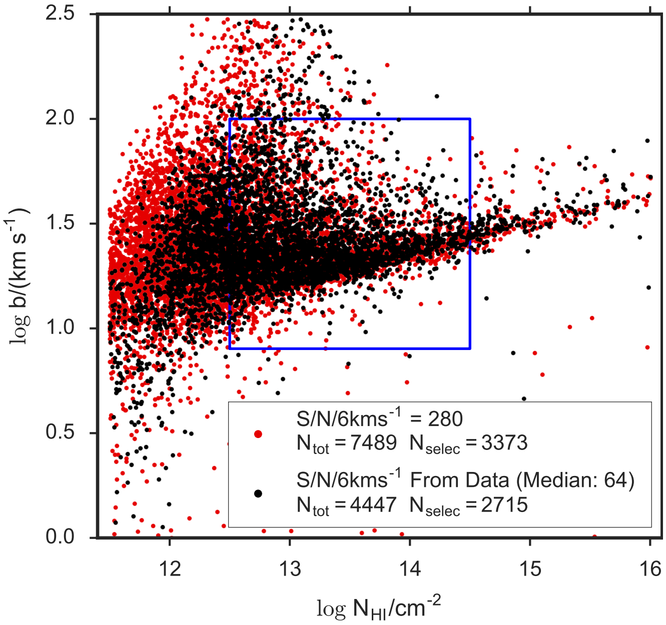


Figure 2.4. Simulated b-N_{HI} distributions at $z = 2.4$ with different SNR applied to lines of sight. The b-N_{HI} distributions were generated by Voigt profile fitting the same 80 mock skewers from collisionless simulations and adding noise and resolution effects. The b-N_{HI} distribution based on high SNR skewers (red) has a higher number of fitted lines than the distribution created based on the SNR distribution of our data at this redshift (black). The high SNR distribution is more complete at low $\log N_{\mathrm{HI}}$ and high $\log b$. The blue box shows the region chosen for our further analysis. The completeness is comparable within this box. The thermal parameters used in these mocks are $\gamma = 1.5$, $\log T_0/\mathrm{K} = 4.04$ and the smoothing length is $\lambda_P = 47\,\mathrm{kpc}$. Figure credit: Hiss et al. (2018).

parameters $8\,\mathrm{km\,s^{-1}} \leq b \leq 100\,\mathrm{km\,s^{-1}}$ both distributions are similarly populated. Therefore, for the cutoff fitting procedure, we will only use the part of the b-N_{HI} distribution with $\log N_{\mathrm{HI}}/\mathrm{cm}^{-2} > 12.5$, which is the convention adopted in Schaye et al. (2000) and Rudie et al. (2012a). We also want to avoid saturated absorbers, i.e. $N_{\mathrm{HI}} > 10^{14.5}\mathrm{cm}^{-2}$, to make sure that we are using only well-constrained column densities. Lines with $b < 8$ km s^{-1} are excluded, because these are most likely metal-line contaminants or VPFIT artifacts. Lines with $b > 100$ km s^{-1} are excluded as well, because the turbulent-broadening component dominates over thermal broadening for such broad lines. This is the same convention used in Rudie et al. (2012a) and is shown as a blue box.

Additionally, we decided, based on Schaye et al. (1999), to exclude points that have relative errors worse than 50% in b or N_{HI}. This is done to avoid using weakly constrained absorption line parameters in the procedure, as they lie mostly in the part of the b-$N_{\mathrm{H\,I}}$ distribution that is affected by the SNR effects described in Figure 2.4 at high b and low $N_{\mathrm{H\,I}}$. Lines with $b < 11\mathrm{km\,s^{-1}}$, i.e. below the low-$b$ envelope of the distribution, are generally not excluded by this procedure.

2.1.4. Metal Masking

It is well known that narrow absorption lines arising from ionic metal-line transitions contaminate the Lyα forest and will particularly impact the lower $b \lesssim 10\,\mathrm{km\,s^{-1}}$ region of the b-$N_{\mathrm{H\,I}}$ distribution if treated as Lyα absorption, thus possibly making the determination of the position of the lower envelope of the b-$N_{\mathrm{H\,I}}$ distribution ambiguous. To address this issue, we remove lines from our sample that are potentially of metal origin.

However, narrow absorption lines are not necessarily metal-line contaminants. We visually inspected the absorption lines with $b \lesssim 10\mathrm{km\,s^{-1}}$ in every sightline and found that although many could be identified as metal lines wrongly fit as Lyα absorption, a comparable number are simply narrow components that VPFIT adds to obtain the best fit to complex Lyα absorption features. The latter are a property of the fitting procedure and should not be excluded, as they are present in both data and the simulated spectra that we use to conduct our analysis[5]. In order to diminish the problem of metal-line contamination we remove metal-line contaminants combining automated and visual identification methods, which we describe in detail below.

Metals are typically associated with strong H I absorption, or they can be identified via associations with other ionic metal line transitions. Therefore, we identified DLAs based on the damping wings of the absorption profiles and determined their redshifts with the help of associated metal absorption redward of the Lyα emission peak of the QSO in question. The redshifts of other strong metal-absorption systems not associated with a DLA within the data coverage or significantly shifted from a DLA are determined by searching for typical doublet absorption systems (mostly Si IV, C IV, Mg II and Al III) redward of the QSO's Lyα emission peak. In both cases, the doublets are identified based on their characteristic $\Delta\lambda$ (see Table 2.3) and line ratios.

Additionally, we selected lines with Doppler parameters $b < 11\ \ \mathrm{km\,s^{-1}} \times (N_{\mathrm{H\,I}}\,/\,10^{12.95}\mathrm{cm^{-2}})^{(1.15-1)}$ in the b-$N_{\mathrm{H\,I}}$ distributions (red line in Figure 2.5) and traced them back to their positions in the spectra. This relation was chosen based on visual inspection of the b-$N_{\mathrm{H\,I}}$ distributions at all redshift bins and chosen to lie underneath the lower envelope of the full dataset. We checked wether we could find a match for different doublet ionic transitions within the Lyα forest for these lines (typically Si IV, C IV

---

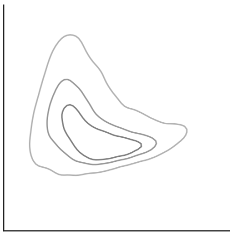
[5] For a discussion about how to circumvent the ambiguities associated with line deblending, see McDonald et al. (2001).

Table 2.3. List of masked metal transitions.

Absorber	$\lambda_{\mathrm{rest}}/\mathrm{Å}$	Absorber	$\lambda_{\mathrm{rest}}/\mathrm{Å}$
O VI[a]	1031.9261	Si IV[a]	1402.770
C II	1036.3367	Si II	1526.7066
O VI	1037.6167	C IV[a]	1548.195
N II	1083.990	C IV[a]	1550.770
Fe III	1122.526	Fe II	1608.4511
Fe II	1144.9379	Al II	1670.7874
Si II	1190.4158	Al III	1854.7164
Si II	1193.2897	Al III	1862.7895
N I	1200.7098	Fe II	2344.214
Si III[a]	1206.500	Fe II	2374.4612
N V	1238.821	Fe II	2382.765
N V	1242.804	Fe II	2586.6500
Si II[a]	1260.4221	Fe II	2600.1729
O I	1302.1685	Mg II	2796.352
Si II	1304.3702	Mg II	2803.531
C II	1334.5323	Mg I	2852.9642
C II*	1335.7077	Ca I	3934.777
Si IV[a]	1393.755	Ca I	3969.591

[a] Strongest transitions. The technique based on high density Lyα systems filters only for these transitions.

and Mg II) by testing for the $\Delta\lambda$ and line ratios. We then confirmed them by finding corresponding absorption of other metals redward of the Lyα emission peak of the QSOs at the same redshift. We then tested if the remaining lines below the lower envelope of the b-$N_{\mathrm{H\,I}}$ distribution were any of the metal transitions listed in Table 2.3 by checking if other metal transitions and Lyα absorption appear at the same redshift. The redshifts of systems positively identified as a metal-line absorption with this method are stored. Candidate metal-line absorbers only identified via a single metal feature or a doubtful doublet feature, i.e. with one of the components possibly within a superposition of absorption features, were not considered as secure metal identification and thus are not masked. Given that it targets the absorbers found during the Voigt profile fitting procedure, this method has the advantage that it allows us to identify metal absorbers within the Lyα forest region.



To further refine our metal-line search, we used a semiautomated procedure to identify high column density ($N_{\rm HI}/{\rm cm}^{-2}$ $\simeq$15) H I absorbers in our sample,[6] as these might also be associated with strong metal absorption. This algorithm identifies groups of pixels in a spectrum that have flux at the relative positions of Lyα, β, γ and higher orders (if available) within a 1σ threshold of zero. The detected systems are then visually compared to theoretical line profiles of absorbers with $\log(N_{\rm HI}/{\rm cm}^{-2}) = 15,\ 16,\ 17$ in Lyα and higher transitions up to Lyγ. If the absorption profile resembles that of a strong absorber, the redshift of the absorption system is saved. If the absorption was stronger than the $\log(N_{\rm HI}/{\rm cm}^{-2}) = 15$ profile, then associated metals were masked (not the H I absorption).

Once we have the redshifts of the metal-absorption systems, we create a mask based on the relative wavelength positions of the metal transitions listed in Table 2.3. All listed transitions are used for generating masks, except for the systems identified with the automated method, i.e. the ones associated with $\log(N_{\rm HI}/{\rm cm}^{-2}) \geq 15$. In this case, we opted for a reduced list of strong ionic transitions (indicated in Table 2.3). In case the position of any line from the `VPFIT` output falls within ±30km s^{-1} of a potential metal line, it is removed from the line list. Additionally, Galactic Ca II (3968, 3933Å) absorption was masked with a ± 150 km s^{-1} window.

Figure 2.5 shows normalized contours for all lines rejected using the narrow-line rejection method described above (gray contour lines) and the lines that were kept (red filled contours) in our sample. We also show the fraction of points rejected in different regions of the b-$N_{\rm H\,I}$ distribution. Our metal line filtering approach will inevitably also filter out lines that are genuine Lyα absorption because of the window size of 30 km s^{-1} used in the narrow-line rejection, removing 24% of the absorbers that are not narrow. This effect is visible in the overlap of rejected and accepted absorbers at $\log b > 11$ km s^{-1}. Nevertheless, we identified and removed 65% of all absorbers in our dataset that are likely to be metal-line contamination within our cutoff fitting range.

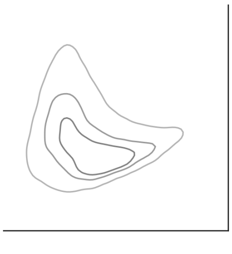
2.1.5. Narrow-Line Rejection

Even after a careful metal-line masking procedure, many unidentified narrow lines still remain in our line lists. These are narrow lines in blends and unidentified metal lines.

One option to avoid these lines is by simultaneously fitting absorption profiles in the Lyβ (or higher transitions) forest, as in Rudie et al. (2012a). While this approach may deliver cleaner b-$N_{\rm H\,I}$ distributions, reproducing the same procedure applied to the data on simulations is very complicated, as it requires modeling of higher-order Lyman series absorption as well. Furthermore, the Rudie et al. (2012a) selection of lines was not completely automated, and decisions about what lines to keep were made by

---

[6] This algorithm was written and tested by John O'Meara.



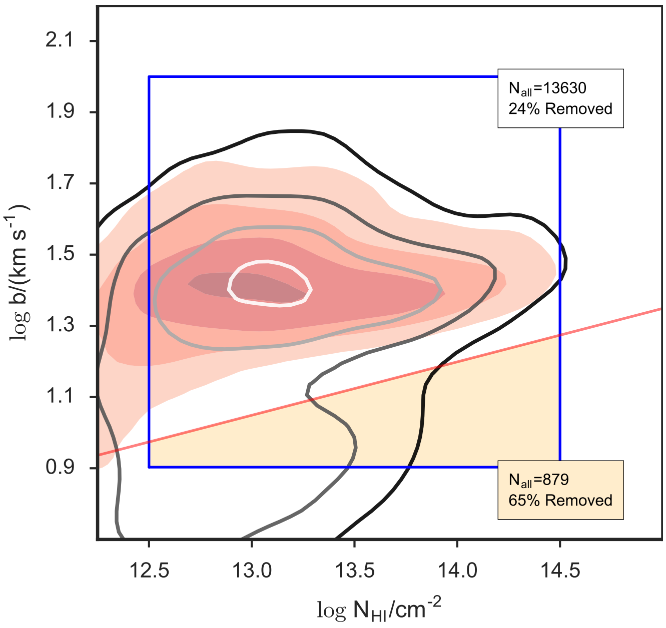


Figure 2.5. All lines in our QSO sample divided into two groups: the ones that were rejected using our narrow-line rejection methods (gray contour lines) and the ones that were kept and will be used for further analysis (red filled contours). For the sake of visibility, we plot the two clouds of points as continuous and normalized density distributions, calculated using a kernel density estimation method. The lines correspond to five equally spaced bins in density, i.e. the 80th, 60th, 40th and 20th percentiles of the corresponding density distributions. The blue square corresponds to our cutoff fitting region. The solid red line broadly represents the dividing line between the bulk of the distributions of broad and narrow lines with with $b < 11\,{\rm km\,s^{-1}} \times (N_{\rm H\,I}/10^{12.95}{\rm cm^{-2}})^{(1.15-1)}$. The orange area illustrates the region mostly affected by narrow lines in our cutoff fitting procedure. The fact that the red contours have little density below the red line indicates that our metal-rejection methods exclude most of the contamination. This happens at the cost of fraction of the usable data, i.e. the lines in the gray contours that are not narrow. The total (both rejected and accepted together) number of lines $N_{\rm all}$ within the blue square is shown above and below the solid red line, as well as the percentage of these lines that were rejected as possible metal absorbers. Figure credit: Hiss et al. (2018).

eye, which cannot be automatically applied to simulations (see Rudie et al. 2012b, for more details). Therefore, in this work, we chose to use only the Lyα forest region.

Since there is no obvious way of filtering the remaining narrow lines, we need to come up with a rejection mechanism to filter them and diminish their impact on our cutoff fitting procedure. To account for this problem, Schaye et al. (1999) removed all

the points in the b-$N_{\mathrm{H\,I}}$ distribution where the best-fitting Hui-Rutledge function[7] to the b-distribution dropped below 10^{-4} at the low-b end. In Rudie et al. (2012a), the authors applied a more sophisticated algorithm that iteratively removes points from the b distributions (with $b < 40$ km s^{-1}) in $\log N_{\mathrm{H\,I}}$ bins in case they are more than 2σ away from the mean.

In this chapter, we approach this problem in a very similar way as in Rudie et al. (2012a) Our rejection algorithm bins the points within $12.5 \leq \log(N_{\mathrm{H\,I}}/\mathrm{cm}^{-2}) \leq 14.5$ into six bins of equal size in $\log(N_{\mathrm{H\,I}}/\mathrm{cm}^{-2})$. Only points with $b < 45$ km s^{-1} are [8] used for the 2σ rejection process. For each of the column-density bins, we compute the mean and the variance of b. Points below 2σ of the mean are excluded. This procedure is iterated until no points are excluded. Finally, after the last iteration, we fit a line to the $\log b_{2\sigma}$ values of each $\log(N_{\mathrm{H\,I}})$ bin. Once the position of this line is determined, we exclude all points below it from the original sample. We have tested this algorithm for the effect of varying the σ threshold and found that the end results are consistent with each other within the errors.

In Figure 2.6, we show a histogram with the number of absorbers in every redshift bin of our data sample and the effects of rejections. Here we see that the 2σ rejection excludes a relatively small fraction of the points in the b-$N_{\mathrm{H\,I}}$ distribution.

2.1.6. Fitting the Cutoff in the b-N_{HI} Distribution

Once we have the b-$N_{\mathrm{H\,I}}$ distributions, we want to determine where the thermal state sensitive cutoff is positioned. The position of the cutoff is calculated using our version of an iterative fitting procedure first introduced by Schaye et al. (1999) and also used in Rudie et al. (2012a). The function used for the cutoff of the b-$N_{\mathrm{H\,I}}$ distribution is given by

$$\log b_{\mathrm{th}} = \log b_0 + (\Gamma - 1)\log(N_{\mathrm{H\,I}}/N_{\mathrm{HI,0}}). \tag{2.2}$$

where b_0 is the minimal broadening value at column density $N_{\mathrm{HI,0}}$ and Γ is the index of this power-law relation.

Although the value of $N_{\mathrm{HI,0}}$ is essentially just a normalization, as we will motivate further in our discussion of the estimation of $N_{\mathrm{HI,0}}$ in § 2.3.2, it is convenient to choose it so that it corresponds to the column density of a typical absorber at the mean density

---

[7] A one parameter function that describes the distribution of Doppler parameters b under the assumptions that $\ln\tau$ is a Gaussian random variable, where τ is the optical depth, and that absorption lines arise from peaks in the optical depth (Hui & Rutledge 1999).

[8] The cut in $b < 45$ km s^{-1} was chosen to be higher than the one used in Rudie et al. (2012a), because lower values were causing the rejection at 2σ to lie too close to the estimated position of the cutoff at some of the redshift bins. The higher cut in b increases the dispersion per bin, making our rejection more conservative.

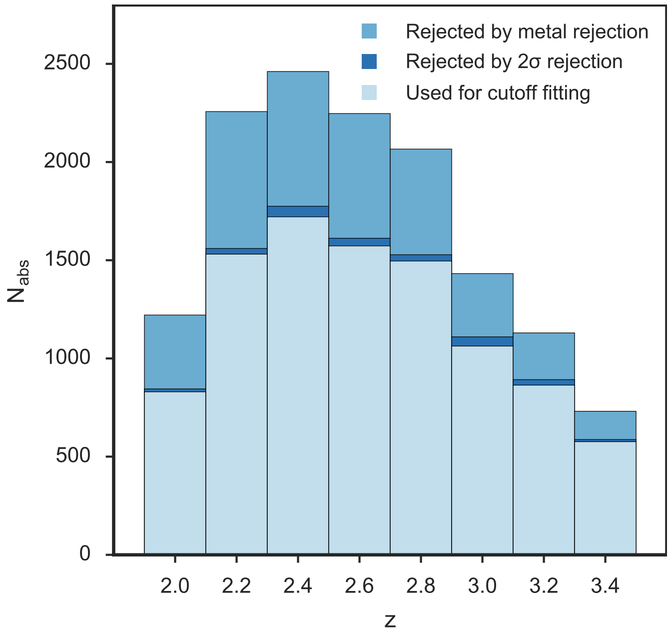


Figure 2.6. Number of absorbers identified by `VPFIT` per redshift bin. The histogram shows the number of lines within the cutoff fitting range after metal lines rejection and the 2σ rejection were applied. Figure credit: Hiss et al. (2018).

of the IGM. Schaye (2001) showed that an absorber corresponding to an overdensity $\Delta = \rho/\rho_0$ with size of order of the IGM Jeans scale will have a column density

$$N_{\rm HI} \simeq 10^{13.23} {\rm cm}^{-2} \Delta^{3/2} \frac{T_4^{-0.22}}{\Gamma_{\rm HI,12}} \left(\frac{1+z}{3.4}\right)^{9/2}, \tag{2.3}$$

where $\Gamma_{\rm HI}$ is the photoionization rate of H I (in units of $10^{-12}\ {\rm s}^{-1}$) and T_4 is the temperature of the absorbing gas in units of $10^4\,K$. We compute $N_{\rm HI,0} = N_{\rm HI}(\Delta = 1)$ at each redshift using this eqn. and discuss how it impacts our calibration in § 2.3.2.

In our iterative cutoff fitting procedure, we fit eqn. 2.2 to points in the b-$N_{\rm HI}$ distribution using a least squares (LS) minimization algorithm that takes into account the errors reported by `VPFIT`. Note that previous works (Schaye et al. 1999; Bolton et al. 2014; Rorai et al. 2018) have used an least absolute deviation (LD) minimization method for fitting. For a method comparison and discussion, see § 2.4.4.

The first step of the cutoff fitting procedure is to fit eqn. 2.2 to all points that are within $10^{12.5}{\rm cm}^{-2} < N_{\rm HI} < 10^{14.5}{\rm cm}^{-2}$ and $8\ {\rm km\,s}^{-1} < b < 100\ {\rm km\,s}^{-1}$. The first iteration results in a fit that falls somewhere close to the mean of the distribution. Then we compute the mean absolute deviation in terms of $\log b$ of all N absorbers with

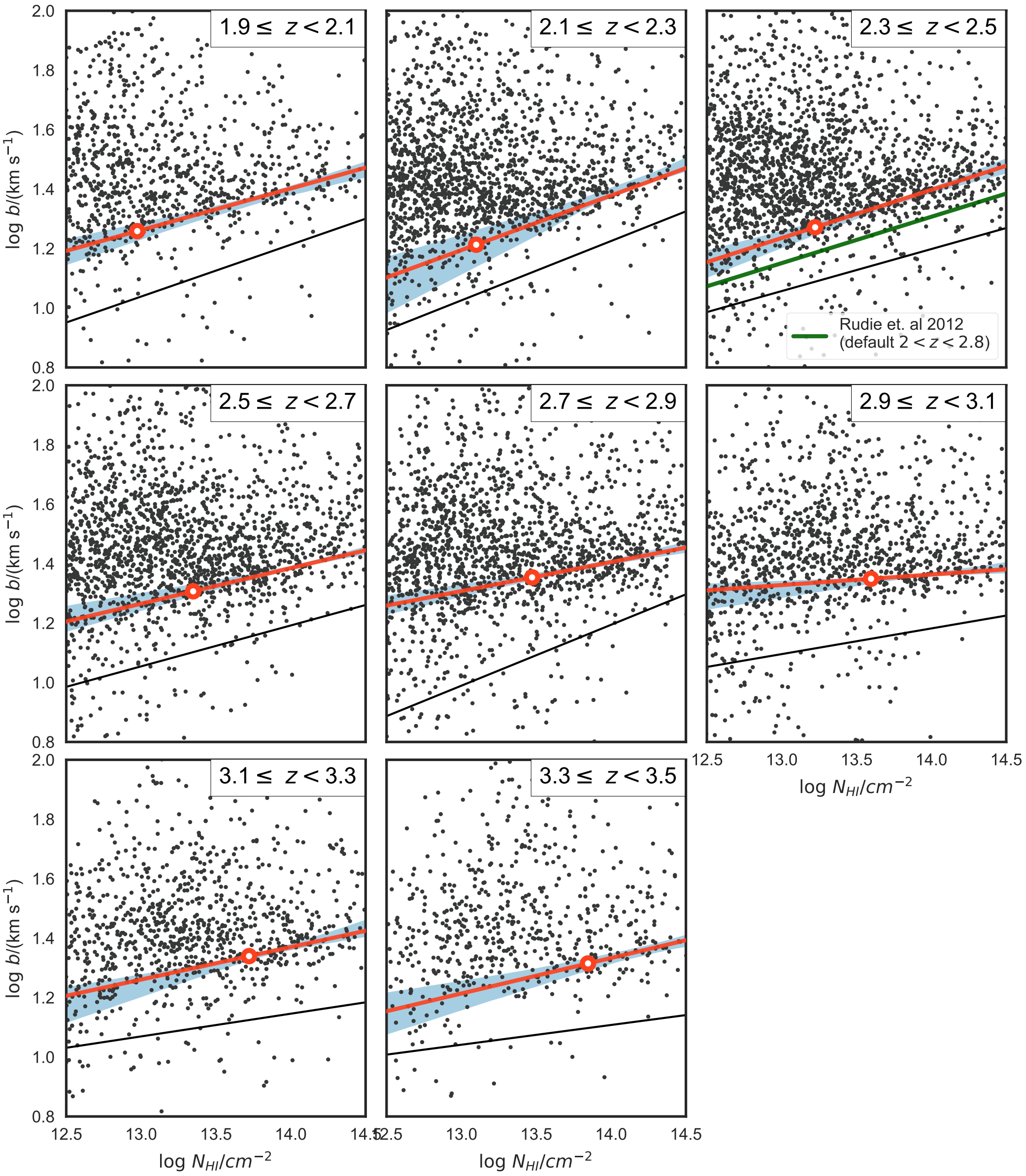


Figure 2.7. The b-$N_{\rm HI}$ distributions in the redshift range $1.9 \leq z < 3.5$ in $\delta z = 0.2$ bins (corresponding to the sightlines in Figure 2.2). The best cutoff fits (red) and 2σ-rejection (black) lines are overplotted. The shaded blue region represents the 68% confidence region of the fits to bootstrap realizations at every column density. The corresponding $N_{\rm HI,0}$ is plotted as an open red point and is calculated by plugging in the bin center redshift into eqn. 2.12. These measurements allow us to access the evolution of b_0 and Γ as a function of redshift. Figure credit: Hiss et al. (2018).

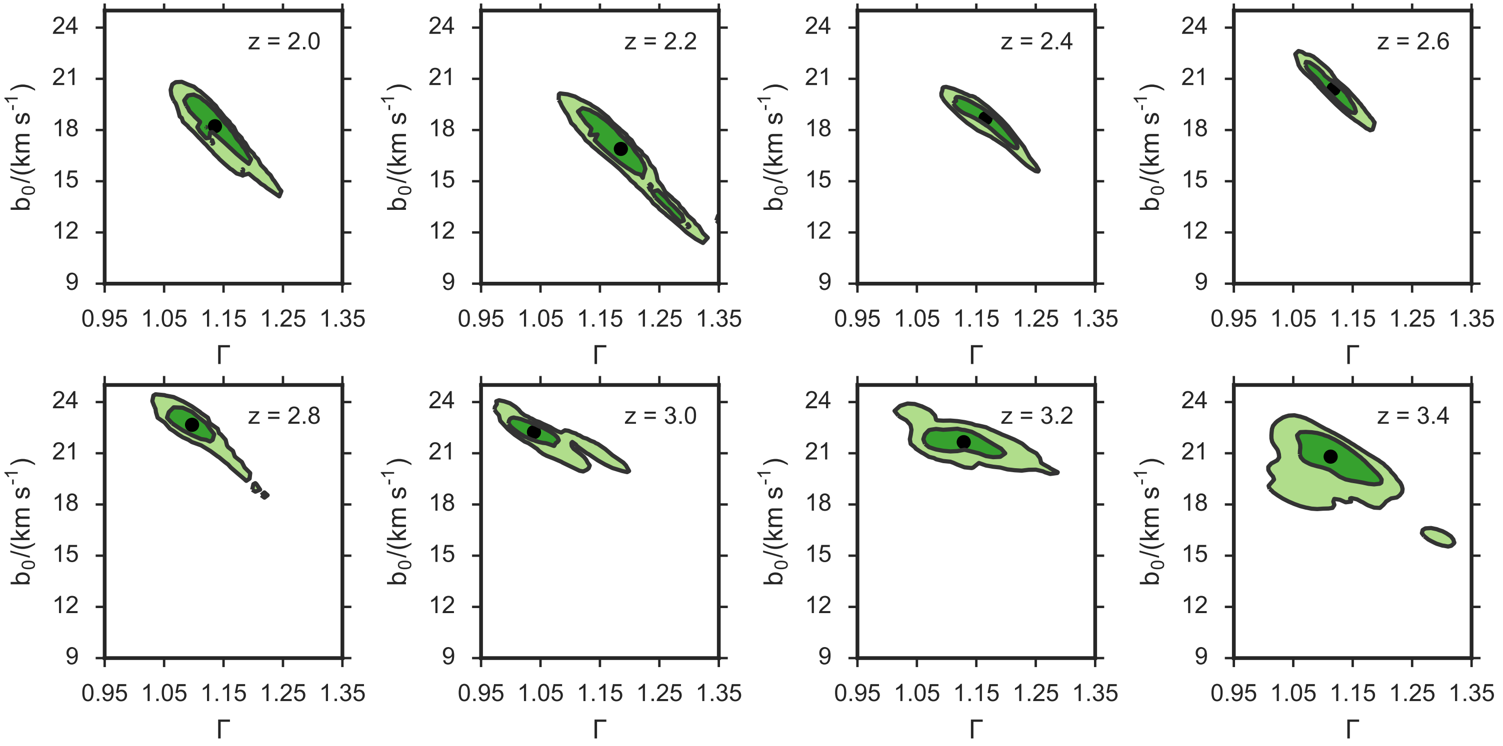


Figure 2.8. Resulting $p\,(b_0, \Gamma)$ PDFs generated from cutoff fits to 2000 bootstrap realizations of the b-$N_{\mathrm{H\,I}}$ distributions at each redshift. The 68% confidence levels are plotted in dark green and 95% in light green. The black points correspond to the median of the marginal distributions of b_0 and Γ.

respect to the first fit:

$$\langle|\delta \log b|\rangle = \frac{1}{N}\sum_{i}^{N} |\log b_i - \log b_{\text{th}}(N_{\text{H I},i})|. \tag{2.4}$$

Notice that this takes the deviations both above and below the fit into account. All the points that have a Doppler parameter with $\log b > \log b_{\text{th}} + \langle|\delta \log b|\rangle$ are excluded in the next iteration. This process is repeated without the points excluded in the previous iteration until no points are more than one absolute mean deviation above the fit, which defines convergence. After convergence, the absorbers that are more than one mean deviation below the last fit are excluded. The remaining points are used for the final fit.

2.1.7. Data Cutoff Fitting Results

Figure 2.7 shows the $\log b(z)$-$\log N_{\text{H I}}(z)$ distributions resulting from the Voigt profile fitting procedure and the respective cutoff fits (red) and 2σ rejection lines (black). The values of $N_{\text{HI},0}$ chosen for each cutoff fit are calculated using eqn. 2.12 at the central redshift of each bin. Their values are plotted as open red circles. We determine the uncertainty in the cutoff fit parameters via a bootstrap procedure. For this purpose, we generate the PDF $p\,(b_0, \Gamma)$ by bootstrapping the cutoff fitting procedure 2000 times using random realizations of the b-$N_{\text{H I}}$ distribution points with replacement. This results in a list with 2000 pairs of (b_0, Γ). The 68% confidence region of the bootstrap cutoff fits is shown in light blue. For illustration, a kernel density estimation of $p\,(b_0, \Gamma)$ at every redshift is shown in Figure 2.8. The anticorrelation between b_0 and Γ is evident.

2.2. Hydrodynamic Simulations

In this section, we describe how we generate Lyα forest mock spectra from Nyx hydrodynamic simulations (Almgren et al. 2013; Lukić et al. 2015) with different combinations of the underlying thermal parameters T_0, γ and λ_P. We apply the exact same Voigt profile and b-$N_{\text{H I}}$ distribution cutoff fitting algorithms as for the data in order to calibrate the relations between the parameters that describe the cutoff (b_0 and Γ) and the thermal parameters (T_0 and γ) while marginalizing over different values of the pressure smoothing scale λ_P.

The evolution of DM in Nyx is calculated by treating DM particles as self-gravitating Lagrangian particles, while baryons are treated as an ideal gas on a uniform Cartesian grid. Nyx uses a second-order accurate piecewise parabolic method (PPM) to solve for the Eulerian gas dynamics equations, which accurately captures shock waves. For more details on the numerical methods and scaling behavior tests, see Almgren et al.

(2013) and Lukić et al. (2015). These simulations also include the physical processes needed to model the Lyα forest. The gas is assumed to be of primordial composition, with hydrogen and helium contributing 75% and 25% by mass. All relevant atomic cooling processes, as well as ultraviolet (UV) photoheating, are modeled under the assumption of ionization equilibrium. Inverse Compton cooling off the microwave background is also taken into account. We used the updated recombination, cooling, collision ionization, and dielectric recombination rates from Lukić et al. (2015).

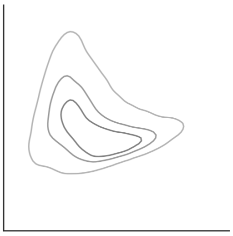
As is standard in hydrodynamic simulations that model the Lyα forest forest, all cells are assumed to be optically thin to radiation. Radiative feedback is accounted for via a spatially uniform but time-varying UVB radiation field, input to the code as a list of photoionization and photoheating rates that vary with redshift (e.g. Katz et al. 1992). We have created a grid of models that explore very different thermal histories combining different methodologies. First, we have used the approach presented in Oñorbe et al. (2017), which allows us to vary the timing and duration of reionization and its associated heat injection, enabling us to simulate a diverse range of reionization histories. This method allows us to create the H I, He I and He II photoionization and photoheating rates, which are inputs to the Nyx code, by volume-averaging the photoionization and energy equations. We direct the reader to Oñorbe et al. (2017) for the details of this method. On top of this, we also use the methodology first introduced by Bryan & Machacek (2000) of rescaling the photoheating rates by factor, A, as well as making the heating depend on density according to Δ^B (Becker et al. 2011), with B also being a free parameter. Combining all these approaches allows us to build a large set of different thermal histories and widely explore the thermal parameter space of T_0, γ and λ_P at different redshifts.

The Thermal History and Evolution in Reionization Models of Absorption Lines (THERMAL)[9] suite consists of more than 60 Nyx hydrodynamic simulations with different thermal histories and $L_{\mathrm{box}} = 20\,\mathrm{Mpc}/h$ and 1024^3 cells based on a Planck Collaboration et al. (2014) cosmology $\Omega_m = 0.3192$, $\Omega_\Lambda = 0.6808$, $\Omega_b = 0.04964$, $h = 0.6704$, $n_s = 0.96$, $\sigma_8 = 0.826$. As shown in Lukić et al. (2015) for a Haardt & Madau (2012) model, simulations of this box size and larger ones result in nearly the same distribution of column densities and Doppler parameters for the range of these parameters used in this study. The suite also has some extra simulations with different cosmological seeds, box size, resolution elements, and/or cosmology to provide a reliable test bench for convergence and systematics associated with different observables. For all simulations, we have data for every $\Delta z = 0.2$ from $z = 6.0$ down to $z = 1.6$, as well as at $z = 1.0,\ 0.5$ and 0.2.

In this chapter, we use a subset of 26 simulations from the THERMAL suite that were selected to optimize the space of the thermal parameters (described below) within the

---

[9] Url: thermal.joseonorbe.com



redshift range in which we are interested: $2.0 \leq z \leq 3.4$. The thermal parameters T_0 and γ are extracted from the simulations by fitting a power-law T-ρ relation to the distribution of gas cells as described in Lukić et al. (2015). In order to determine the pressure smoothing scale λ_P, the cutoff in the power spectrum of the real-space Lyα flux F_{real} is fitted. Here F_{real} is the flux each position in the simulation would have given its temperature and density but neglecting redshift space effects (see Kulkarni et al. 2015).

2.2.1. Skewer Generation

In order to model lines of sight through the IGM, we extract a random subset of hydrogen density skewers from our simulations that run parallel to the box axes. These are transformed into Lyα optical depth skewers (we refer to Lukić et al. 2015 for specific details about these calculations). The corresponding flux skewer F, i.e. a transmission spectrum along the line of sight, is calculated from the optical depth using $F = \exp(-A_r\,\tau)$. Here we introduce a scaling factor A_r used for matching our lines of sight to observed mean flux values. This rescaling of the optical depth accounts for the lack of knowledge of the precise value of the metagalactic ionizing background photoionization rate. To this end, we choose A_r so that we match the mean flux evolution shown in Oñorbe et al. (2017), which is a fit at $0.2 < z < 5.85$ based on measurements of various authors (Fan et al. 2006; Kim et al. 2007; Faucher-Giguère et al. 2008a; Becker et al. 2013). Given the extremely high precision with which the mean flux has been measured by these authors, we do not consider the impact of uncertainties in the rescaling value A_r. A discussion of the effects of mean flux rescaling in the models on our results is presented in the Appendix A.2.

2.2.2. Thermal Parameter Grid

We used simulation snapshots at eight different redshifts from $z = 2.0$ to 3.4 in $\delta z = 0.2$ steps, which matches the redshift distribution of our data. We then generated 150 skewers for $2.0 \leq z \leq 3.0$ and 75 skewers for $3.2 \leq z \leq 3.4$[10] for each of the 26 combinations of thermal parameters (T_0, γ and λ_P). Figure 2.9 shows the distribution of thermal parameters chosen. We chose to model the thermal parameters on an irregular grid covering the range $47\,\mathrm{kpc} < \lambda_P < 120\,\mathrm{kpc}$, which is well within the range of measurements by Rorai et al. (2013, 2017b) of $40\,\mathrm{kpc} < \lambda_P < 130\,\mathrm{kpc}$ for $2 < z < 3.6$. For this comparison, we scaled the measurements of Rorai et al. (2013, 2017b) to match

---

[10] These numbers of skewers were chosen based on the computation time needed for Voigt profile fitting $z > 3.2$ mock spectra at high SNR. Adopting these numbers results in nearly the same amount of absorbers in the b-N_{HI} distribution used for cutoff fitting as in our data bins from $z = 2$ to 2.6 and ~ 2500 absorbers from $z = 2.8$ to 3.4.

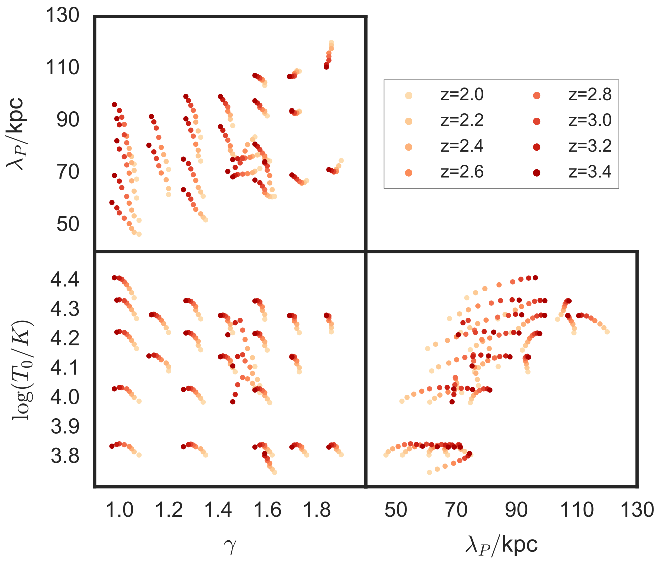


Figure 2.9. Combinations of parameters T_0, γ and λ_P used for generating model skewers used in the calibration process. This grid was generated at redshifts 2.0 to 3.4 in $\delta z = 0.2$ steps to match our data. The evolution of the grid with redshift reflects the thermal history of the Nyx simulations chosen. Figure credit: Hiss et al. (2018).

λ_P as defined in Kulkarni et al. (2015). The grid of parameters of the temperature-density relation covers $0.97 < \gamma < 1.9$ and $5600\ \mathrm{K} < T_0 < 25700\ \mathrm{K}$.

2.2.3. Forward-modeling Noise and Resolution

To create mock spectra, we add the effects of resolution and noise, both based on our data, to our simulated skewers. We model instrumental resolution by convolving the simulated lines of sight with a Gaussian with FWHM = 6 $\mathrm{km\,s^{-1}}$ which is our typical spectral resolution, and rebinning to 3 $\mathrm{km\,s^{-1}}$ pixels afterward. To make our mock spectra comparable to the data, we added noise to the flux based on the error distribution as provided by the data reduction pipelines. First, a random Lyα forest at the same redshift interval is chosen from our QSO sample. A Gaussian PDF is constructed based on the median and a rank-based estimate of the standard deviation of the error distribution of the chosen data segment. Then, for every pixel i in the skewer, we draw random errors ϵ_i from this PDF. We rescale the errors so that $\epsilon_{i,r} = \sqrt{\Delta\lambda_{\mathrm{data}}/\Delta\lambda_{\mathrm{skewer}}} \times \epsilon_i$, where $\Delta\lambda$ is the median wavelength distance between pixels. This accounts for the difference in sampling between data and skewers. Finally, we add a random deviate to the flux F_i drawn from a normal distribution with $\sigma = \epsilon_{i,r}$, which is the error bar attributed to the flux. We do not account for metal-line contaminants in our mock

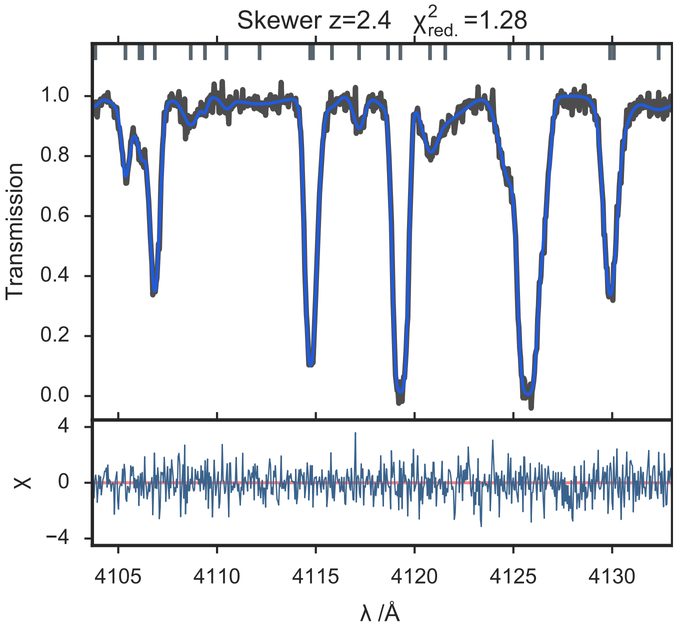


Figure 2.10. Voigt profile fitted simulated line of sight at $z = 2.4$. The blue line is the Voigt profile fitted spectrum, while the black line is the forward modeled skewer. Noise was generated based on our data to achieve a SNR of 52 per pixel at continuum level. The simulation used had a best-fit temperature-density relation with γ=1.52, $\log T_0/\mathrm{K} = 4.07$ and a pressure smoothing scale of $\lambda_P = 70\,\mathrm{kpc}$. Underneath, we plot the corresponding $\chi = (F_{\mathrm{spec}} - F_{\mathrm{fit}})/\sigma_{F_{\mathrm{fit}}}$. Figure credit: Hiss et al. (2018).

spectra, as these are explicitly masked in our data (see § 2.1.4).

2.2.4. Voigt Profile Fitting Simulations

We apply the exact same Voigt profile fitting scheme described in § 2.1.2 to the forward-modeled simulated skewers generated for different combinations of T_0, γ and filtering scale λ_P. A Voigt profile fit of a mock spectrum is shown in Figure 2.10. We then generate a b-$N_{\mathrm{H\,I}}$ distribution for all our models and apply the same cutoff fitting algorithm described in § 2.1.6. We have checked for the effect of applying the 2σ rejection algorithm (as described in § 2.1.5) to b-$N_{\mathrm{H\,I}}$ distributions from simulated spectra and found that, given that there are a few outliers and no metal contamination, the effect is negligible. Therefore, we decided not to apply the 2σ rejection algorithm to simulated b-$N_{\mathrm{H\,I}}$ distributions.

In Figures 2.11 and 2.12, we compare the b-$N_{\mathrm{H\,I}}$ distributions and the respective cutoff fits of the observational data (with metal lines excluded; see section 2.1.4) and mock spectra at all redshift bins. In both data and simulations, a cutoff in the distribution

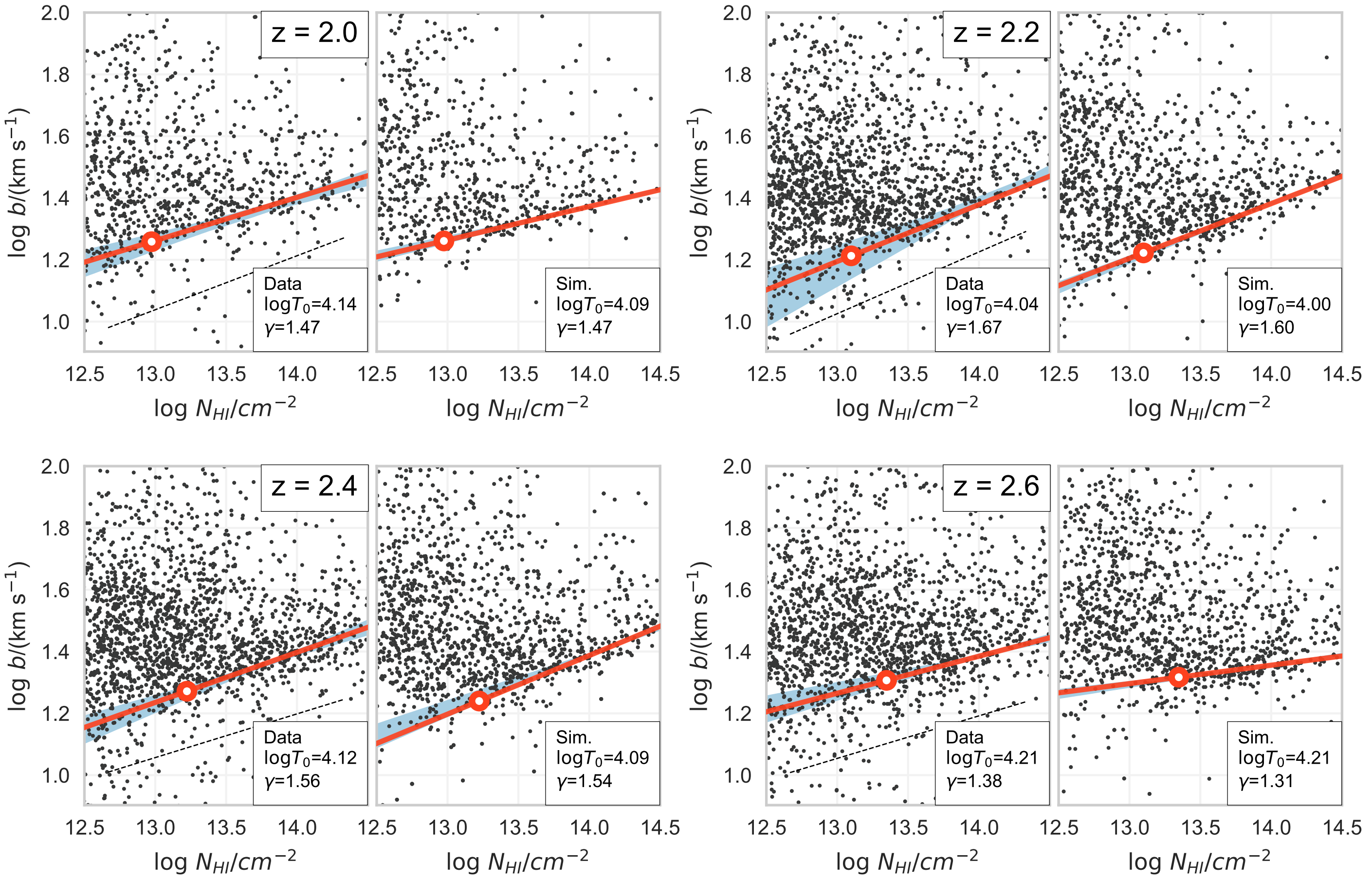


Figure 2.11. Comparison of b-$N_{\rm HI}$ distributions and cutoff fits for the redshift bins $z = 2$, 2.2, 2.4 and 2.6. At each redshift, our data are shown in the left panel, while simulations are shown in the right panel. The simulated distributions shown are the ones that have T_0 and γ closest to our final results (introduced in § 2.4). The best cutoff fits (red) and 2σ rejection (black dashed, data only) lines are overplotted. The light blue region represents the 68% confidence region of the fits to bootstrap realizations at every $N_{\rm HI}$. The value of $N_{\rm HI,0}(z)$ is plotted as an open red circle ($N_{\rm HI,0}$ is motivated in § 2.3.2). The cutoff fitting algorithm responds similarly to data and models once the contamination in the data is removed using the 2σ rejection algorithm. The remaining contamination in the data is still more severe than in the models. This affects how the cutoff fitting procedure reacts to different bootstrap realizations.

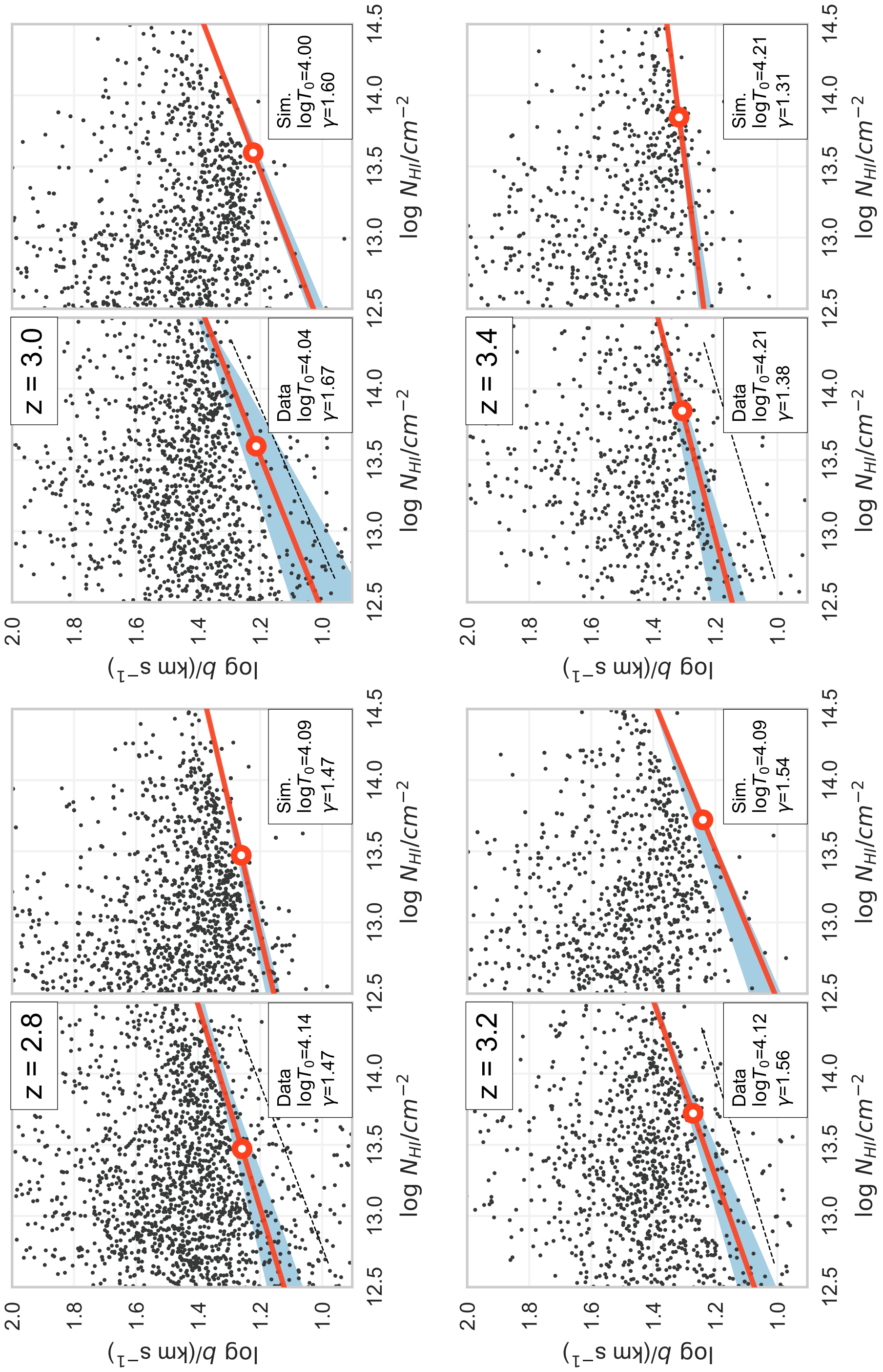


Figure 2.12. Same as in Figure 2.11 for the redshift bins $z = 2.8$, 3.0, 3.2 and 3.4

is evident. We also overplot the best-fit cutoff (red) and the 68% confidence regions (light blue) determined by bootstrapped fits, as described in § 2.1.7. To illustrate the similarities of the data and models, the model shown at each redshift is one that has T_0 and γ closest to our final measurement presented in § 2.4.

The main difference is that the b-$N_{\mathrm{H\,I}}$ distribution of the data exhibits more lines underneath the cutoff, i.e. in the low-b and low-$N_{\mathrm{H\,I}}$ part of the panels in Figures 2.11 and 2.12. As the SNR distribution is comparable in both diagrams, as well as the amount of blended absorption systems, we conclude that, if the model assumptions are right, these are most likely metal lines wrongly identified as Lyα absorption lines. Most of these narrow lines are excluded using the 2σ rejection described in § 2.1.5, as indicated by the black dashed lines in the left panels of Figures 2.11 and 2.12. This leads to the conclusion that we are able to generate b-$N_{\mathrm{H\,I}}$ distributions from our simulations that are similar to those retrieved from data in terms of the cutoff.

2.3. Calibration of the Cutoff Measurements

In this section, we want to use our simulations to quantify how our cutoff observables b_0 and Γ are related to the thermal parameters T_0 and γ. Once this calibration is known, it can be applied to our data and, under the assumption that simulated and measured b-$N_{\mathrm{H\,I}}$ distributions are similar, we can retrieve T_0 and γ from the data.

2.3.1. Formalism

To motivate this calibration, we start with the temperature-density relation (Hui & Gnedin 1997; McQuinn & Upton Sanderbeck 2016), which states that the temperature distribution as a function of gas density is set by the temperature at mean density $T_0 = T(\rho_0)$ and the index γ:

$$\log T = \log T_0 + (\gamma - 1)\log(\rho/\rho_0) \tag{2.5}$$

where γ adjusts the contrast level of how much overdensities are hotter/cooler than underdensities.

In order to construct a relation between b_0 and T_0 as well as between Γ and γ we follow the Ansatz presented by Schaye (2001). It states that the overdensity (ρ/ρ_0) and the overdensity in terms of the column density ($N_{\mathrm{H\,I}}/N_{\mathrm{HI},0}$), where $N_{\mathrm{HI},0}$ is the column density corresponding to the mean density ρ_0, are connected via a power law

$$\log(\rho/\rho_0) = A + B\log(N_{\mathrm{H\,I}}/N_{\mathrm{HI},0}). \tag{2.6}$$

Furthermore, for absorbers along the cutoff for which turbulent line broadening is negligible, the line broadening is purely thermal, resulting in power-law relation between

b_{th} and T

$$\log T = C + D \log b_{\mathrm{th}}, \tag{2.7}$$

where b_{th} is the thermal Doppler broadening. Combining eqns. (2.5)-(2.6) and (2.7) results in a power-law relation between b_{th} and N_{HI} (eqn. 2.2), which is the functional form that we fit to the cutoff of the b-$N_{\mathrm{H\,I}}$ distribution. The coefficients in eqn. 2.2 can be written as

$$\log b_0 = \frac{1}{D}(\log T_0 - C + A(\gamma - 1)) \tag{2.8}$$

$$(\Gamma - 1) = \frac{B}{D}(\gamma - 1). \tag{2.9}$$

Eqn. 2.2 represents the line of minimal broadening at a given column density $N_{\mathrm{H\,I}}$ (therefore b_{th}), because absorbers in this relation are strictly thermally broadened. If the normalization constant $N_{\mathrm{H\,I},0}$ is chosen so that it represents the column density value of a cloud with mean density, then $A = 0$ (see eqn. 2.6), i.e. the dependency on γ disappears from $\log b_0$ in eqn. 2.8. Taking this into account and redefining $\kappa = \frac{D}{B}$ we can rewrite these equations as:

$$\log T_0 = D \log b_0 + C \tag{2.10}$$

$$(\gamma - 1) = \kappa(\Gamma - 1) \tag{2.11}$$

We can calibrate these relations by fitting the cutoff of mock datasets extracted from our simulations in combination with the same cutoff fitting algorithm we applied to the data. This approach has the advantage that it does not require the assumption that gas is only thermally broadened. Thus, we can account for the effects of pressure smoothing and thermal broadening on the position of the cutoff in a generalized way.

2.3.2. Estimation of $N_{\mathrm{HI},0}$

The motivation for normalizing the $N_{\mathrm{H\,I}}$ values with $N_{\mathrm{H\,I},0}$, is that it simplifies the calibration between the b-$N_{\mathrm{H\,I}}$ relation and the T-ρ relation to be a one-to-one mapping between b_0-T_0 and $\gamma - \Gamma$ (equations (2.10) and (2.11)), with the former governed by two parameters (C, D) and the latter governed by a single parameter κ. In other words, any γ dependency is removed from eqn. 2.10.

However, in general, the mapping between Lyα optical depth and density, and hence between $N_{\mathrm{H\,I}}$ and density, depends on the thermal parameters and the metagalactic photoionization rate $\Gamma_{\mathrm{H\,I}}$. This means that in principle, $N_{\mathrm{H\,I}} = N_{\mathrm{H\,I}}(\Gamma_{\mathrm{H\,I}}, T_0, \gamma)$, which can be seen directly from eqn. 2.3, as the temperature is a function of T_0 and γ. This would require determining $N_{\mathrm{H\,I},0}$ for every single thermal model in order to calibrate the simple relations of eqns. 2.10 and 2.11. Luckily, eqn. 2.3 illustrates that the thermal

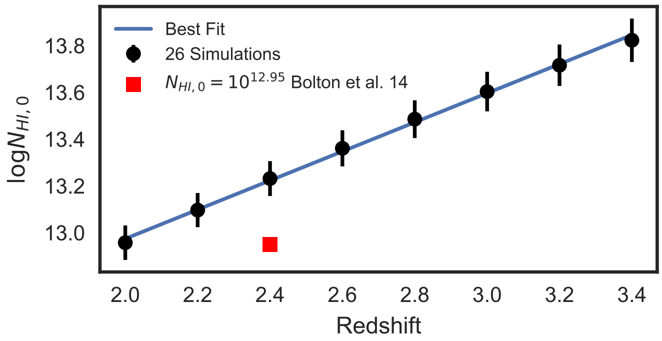


Figure 2.13. Values of $N_{\rm HI,0}$(z) from our simulations. The black points are calculated based on the mean flux correction from Becker et al. (2011) applied to our skewers using eqn. 2.3. The error bars reflect the variance in the mean flux rescaling value (i.e, the strength of the UVB) and T_0 in the 26 models used in this work. The blue line is a linear fit to the black points, which will be used for estimating $N_{\rm HI,0}$(z) in this chapter. For comparison, we show $N_{\rm HI,0}(z = 2.4)$ from Bolton et al. (2014) from hydrodynamic simulations (red square). Figure credit: Hiss et al. (2018).

parameter dependency is quite weak, scaling as $T^{-0.22}$. Instead, the primary dependency is on $\Gamma_{\rm H\,I}$. Furthermore, because one always adjusts the mean UVB to give the same mean flux for different thermal models, the variation of $N_{\rm H\,I}$ with thermal parameters is even further reduced. We will explicitly show that eqn. 2.3 holds in our simulations in the appendix A.4.

The approach that was used in Rudie et al. (2012a) to compute $N_{\rm H\,I,0}$ was to adopt a fixed value of $\Gamma_{\rm H\,I}$ and compute $N_{\rm H\,I}$ analytically, i.e. $N_{\rm H\,I,0} = N_{\rm H\,I}(\Delta = 1)$. Bolton et al. (2014) instead adopted the average value of $N_{\rm H\,I}$ associated with gas at mean density in their simulations. In this chapter, we compute $N_{\rm H\,I,0}$ analytically using eqn. 2.3 evaluated at mean density, i.e. $\Delta = 1$, for the parameters $\Gamma_{\rm H\,I}$ and T_0 from our simulations. Note that we use the effective UVB $\Gamma_{\rm H\,I} = \Gamma_{\rm H\,I,sim}/A_r$, because our simulations were rescaled to give the correct mean flux at a given redshift (see section 2.2.1). Figure 2.13 shows the average and 1σ range of our $N_{\rm H\,I,0}$ values over all of our thermal models as a function of redshift. This confirms that the variation of $N_{\rm H\,I,0}$ over the different thermal models is small, as also argued by Bolton et al. (2014).

Finally, we applied a fit to the mean values of $N_{\rm HI,0}$ over the 26 different simulations, taking the standard deviation as an estimate for the error. The best-fit linear function has the form

$$\log(N_{\rm HI,0}/{\rm cm}^{-2})(z) = a(1+z) + c \tag{2.12}$$

with $a = 0.6225$ and $c = 11.1068$. Throughout this chapter, we will use this function to compute $N_{\rm HI,0}$ values at fixed redshifts.

Our best-fit value of $N_{\rm H\,I,0}$ at $z = 2.4$, $N_{\rm HI,0} \simeq 10^{13.22}$ cm^{-2}, is inconsistent with the

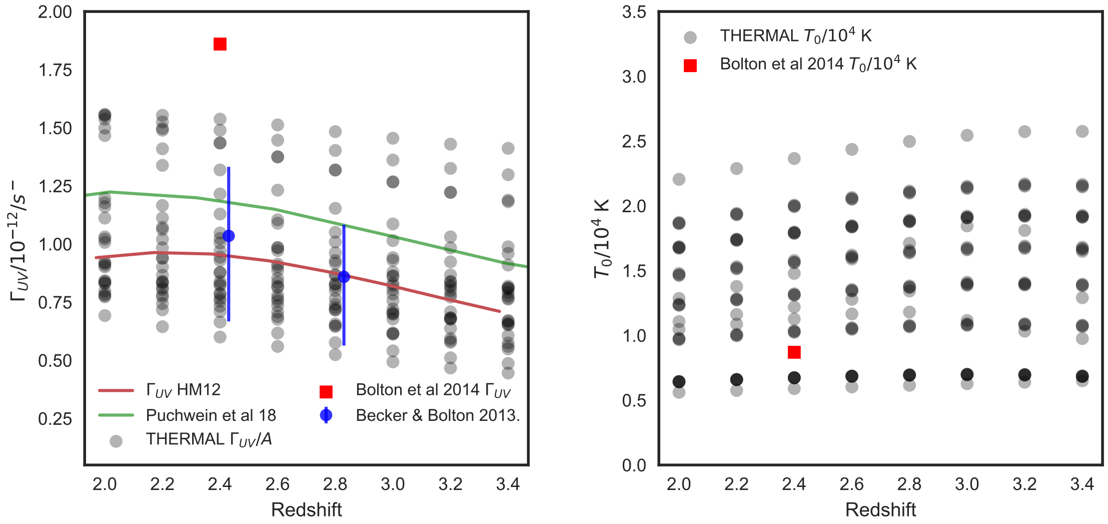


Figure 2.14. The two parameters controlling the relation between column density and density in the IGM (see eqn. 2.3), and therefore the column density at mean density $N_{\mathrm{HI},0}$. **Left:** Comparison of the value of Γ_{HI} at each redshift for the simulations used in our thermal grid to models (Haardt & Madau 2012; Puchwein et al. 2019) and measurement by Becker & Bolton (2013). The value of Γ_{HI} used in Bolton et al. (2014) is shown as a red square. The value chosen in Bolton et al. (2014) is significantly higher than expected and this discrepancy is driving the difference in $N_{\mathrm{HI},0}$ we have with Bolton et al. (2014). **Right:** Comparison of the value of T_0 at each redshift for the simulations used in our thermal grid to T_0 from the simulation in Bolton et al. (2014). This panel illustrates that the values of T_0 are comparable and therefore not driving the difference we see in $N_{\mathrm{HI},0}$.

value measured by Bolton et al. (2014) $N_{\mathrm{HI},0} = 10^{12.95}$ cm^{-2}. This is, see Figure 2.14, presumably because of the high values of Γ_{HI} they needed to match the opacity measurements by Becker & Bolton (2013). Part of this possible discrepancy could be due to the lower temperature in Bolton et al. (2014), but the dependency of N_{HI} on T_0 is too small to drive this difference. While the Bolton et al. (2014) simulations require a value of $\Gamma_{\mathrm{HI}}/10^{-12}\mathrm{s}^{-1} = 1.86$ to match N_{HI} to the optical depth weighted density using eqn. 2.3 (this values were derived from their Figure 1), we use the rescaled values of our simulations, which are consistent with Becker & Bolton (2013) to directly calculate $N_{\mathrm{HI},0}$. This difference of ~ 0.3 dex will certainly lead to inconsistent values of b_0, but since the calibration process is carried out using the same values of $N_{\mathrm{HI},0}$ for both the data and simulations, the calibration will cancel out differences due to $N_{\mathrm{HI},0}$ when dealing with T_0 as long as the scatter due to γ dependency in eqn. 2.10 remains small compared to our statistical error in b_0. We further discuss this in § 2.4.3 when we compare our final measurements to Bolton et al. (2014).

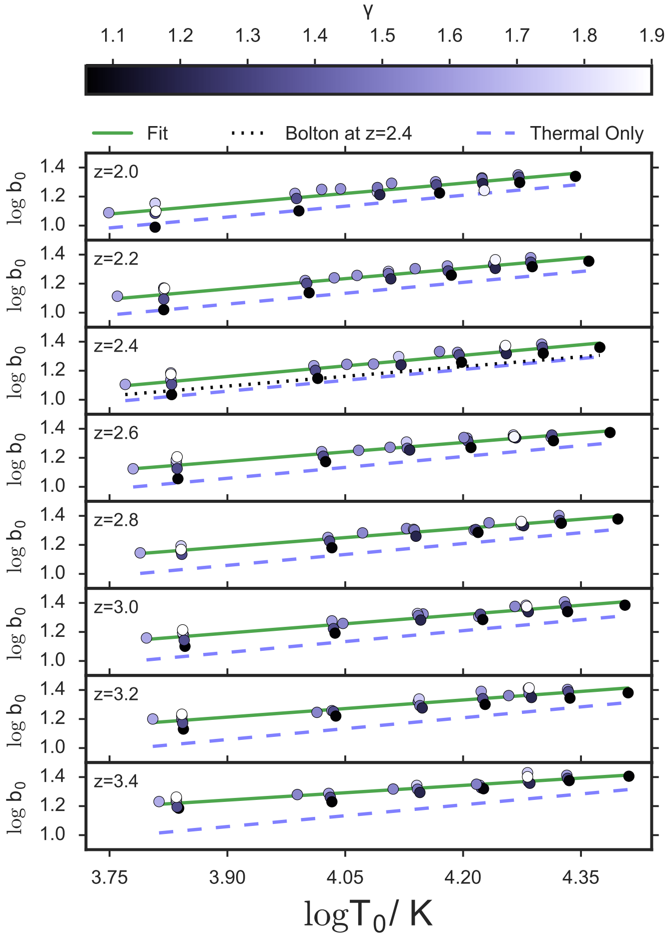


Figure 2.15. Calibration of the $\log b_0$ vs. $\log T_0$ relation. Each point corresponds to a simulated b-N_{HI} distribution. The points are colored based on their γ value. The green lines are the best two-parameter fits to the points. The blue dashed lines represent the case when the value of b_0 is due to pure thermal broadening. The scatter is due to unmodeled λ_P effects, as well as deviations due to γ-dependency of this relation when $N_{\mathrm{HI},0}$ does not exactly correspond to the mean density. At redshift $z = 2.4$ we show the line corresponding to the calibration carried out by Bolton et al. (2014) using hydrodynamic simulations (black dashed). Figure credit: Hiss et al. (2018).

2.3.3. Calibration Using Simulations

In order to generate the calibration between b_0-T_0 and Γ-γ we ran our cutoff fitting algorithm on simulated b-N_{HI} distributions, each constructed from 100 mock spectra drawn from all of our 26 thermal models at each redshift. The results are shown in

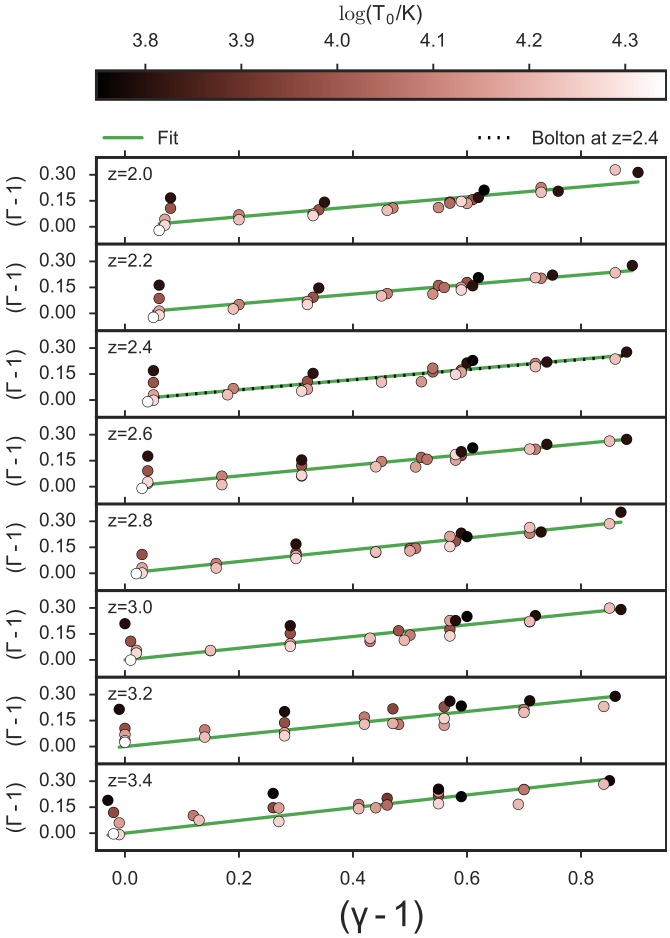


Figure 2.16. Calibration of the $(\Gamma - 1)$ vs. $(\gamma - 1)$ relation. Each point corresponds to a simulated b-N_{HI} distribution. The points are colored based on their T_0 value. The green lines represent the best one-parameter fits to the points. This calibration seems to be independent of the corresponding T_0 and λ_P values. At redshift $z = 2.4$ we show the line corresponding to the calibration carried out by Bolton et al. (2014) using hydrodynamic simulations (black dashed). Figure credit: Hiss et al. (2018).

Figures 2.15 and 2.16, respectively. There we see the simulation input values of T_0 and $(\gamma - 1)$ for our 26 thermal models plotted against the values of b_0 and $(\Gamma - 1)$ extracted from cutoff fits to each b-N_{HI} distribution. Each panel corresponds to a different redshift, which allows us to capture the evolution of the calibration. The green lines are the fits using eqns. (2.10) and (2.11) at every redshift. For comparison, we show the calibration of Bolton et al. (2014) at $z = 2.4$ in black. In the $\log T_0$-$\log b_0$ diagrams, we additionally plot the case in which b_0 arises purely due to thermal broadening, i.e.

$b_0 = \sqrt{2k_B T_0/m_{\mathrm{H\,I}}}$.

The points shown in the diagrams are the median values of b_0 and Γ from 500 random realizations of the b-$N_{\mathrm{H\,I}}$ distributions with replacement rather than the best-fit value of the cutoff parameters of the mock b-$N_{\mathrm{H\,I}}$ distribution. We chose this approach for consistency with how we treated the data, but the results are essentially insensitive to this choice.

Our 26 models have different contributions to the thermal broadening b_0 due to the different values of the pressure smoothing scale λ_P. Similarly, the fact that we assumed one value of $N_{\mathrm{H\,I},0}$ for all models with the same redshift will introduce a small γ dependency in the $\log T_0$-$\log b_0$ relation. We want to include our lack of knowledge about λ_P and additional effects in the calibration by quantifying the amount of scatter that they add into the calibration relations. This is done by simultaneously fitting equations (2.10) and (2.11) to the same 2000 bootstrap realizations of the points in the $\log T_0$-$\log b_0$ and $(\gamma - 1)$-$(\Gamma - 1)$ diagrams with replacement. The best-fit values for every bootstrap realization are stored, giving us the approximated PDFs $p(D, C)$ and $p(\kappa)$. We will quantify to what degree systematic dependencies in our calibration affect our final measurements in the appendix A.3.

For illustration, the calibration values as a function of redshift are shown in Figure 2.17. The error bars correspond to the 68% confidence intervals of $p(\kappa)$ and the marginal distributions of $p(D, C)$. The errors in κ are small because the scatter in the $(\gamma-1)$-$(\Gamma-1)$ relation is only slightly driven by dependencies on T_0 or λ_P.

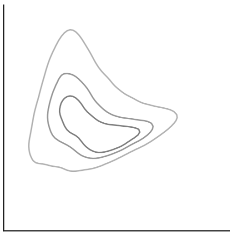
While we agree with the measurements of C, D from Bolton et al. (2014) at $z = 2.4$ in terms of the marginalized distributions of C, D, their calibration values are about 2σ off in terms of the joint PDF $p(C, D)$ as shown in Figure 2.18. This could be attributed to the difference in method used for cutoff fitting (Bolton et al. 2014 uses LD minimization, while we use a LS minimization approach for the cutoff fitting), as well as the difference in $N_{\mathrm{H\,I},0}$. The calibration constant κ between $(\gamma - 1)$ and $(\Gamma - 1)$ we derived agrees within 1σ with the value reported by Bolton et al. (2014).

The impact of the calibration differences is further discussed when we compare our T_0 and γ results to previous works in § 2.4.3.

2.4. Results

2.4.1. Evolution of T_0 and γ

Concerning the evolution of γ, the first conclusion we can draw directly from the data cutoff measurements shown in Figure 2.7 is that a positive $(\Gamma - 1)$ is preferred for all redshift bins. This implies (see eqn. 2.11) that a positive temperature-density relation index $(\gamma - 1)$ is favored at all redshifts probed. In the $p\,(b_0, \Gamma)(z = 3)$ panel in Figure 2.8



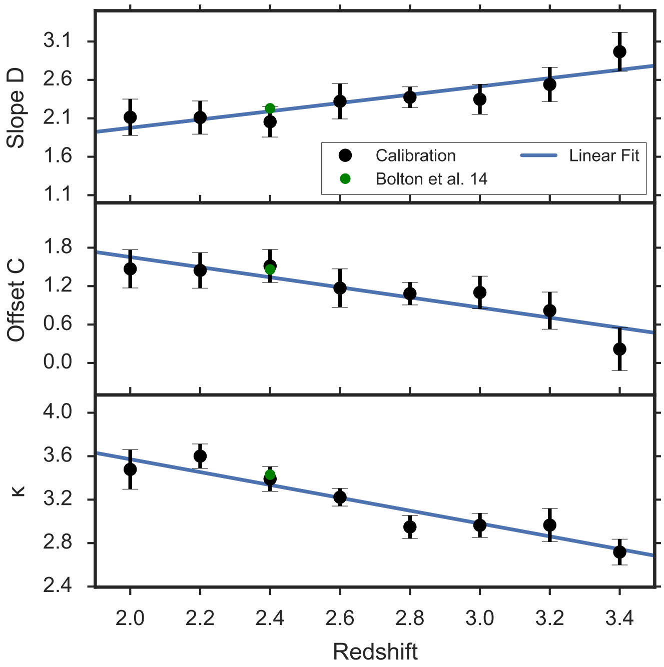


Figure 2.17. Upper panels: bootstrapped fit values to the $\log b_0$ vs. $\log T_0$ relation. The error bars reflect the 68% confidence levels of the marginal distributions of the bootstrapped $p(D, C)$ PDF at each redshift. **Lower panel:** bootstrapped fit values to the $(\Gamma - 1)$ vs. $(\gamma - 1)$ relation. The error bars reflect the 68% confidence levels of the bootstrapped $p(\kappa)$ PDF at each redshift. The blue lines are linear fits to guide the eye. Figure credit: Hiss et al. (2018).

about 4% of the points in $p\,(b_0, \Gamma)$ are consistent with $\Gamma < 1$.

Having both the cutoff measurements and the calibration in hand, we can now estimate T_0 and γ. It is clear from Figure 2.8 that covariance in the cutoff fits will lead to a similar covariance between T_0 and γ, and furthermore, that the scatter in our calibration quantified in Figure 2.17 has to be incorporated into the error budget. To include all of these effects and arrive at the joint probability distribution $p(T_0, \gamma)$ we adopt a Monte Carlo approach as follows. We combine 2000 bootstrapped b_0 and Γ pairs in $p\,(b_0, \Gamma)$ with every one of the 2000 points in the bootstrapped calibration PDFs $p(D, C)$ and $p(\kappa)$ from simulations using eqns. (2.10) and (2.11) at every redshift bin. The contours of the 2000×2000 points in $p(T_0, \gamma)(z)$ estimated via kernel density estimation at every redshift are shown in Figure 2.19. Comparison with Figure 2.8 indicates that the shapes of the T_0-γ contours are qualitatively similar to those of the b_0-Γ contours, which results from noise and degeneracy in fitting the cutoff. This is expected from eqns. 2.10 and 2.11. The contours are slightly broadened by the calibration uncertainty.

Note that uncertainties in T_0 and γ are dominated by the statistical errors of b_0 and Γ due to the high precision of the calibration process.

The evolution of the temperature at mean density T_0 and index of the temperature-density relation γ measured in this chapter is shown in Figures 2.20 and 2.21. The error bars are calculated using the 16th and 84th percentiles of the marginal distributions of T_0 and γ from $p(T_0, \gamma)$. The main features are that the temperature at mean density increases from $z = 3.4$ to $z = 2.8$ (peaking at $T_0 \simeq 20000$ K), while γ has its lowest value, $\gamma = 1.12$, at $z = 3.0$. From $z = 2.8$ to 2.0, T_0 decreases again toward $T_0 \simeq 10000$ K, while γ increases gradually toward $\gamma \simeq 1.6$.

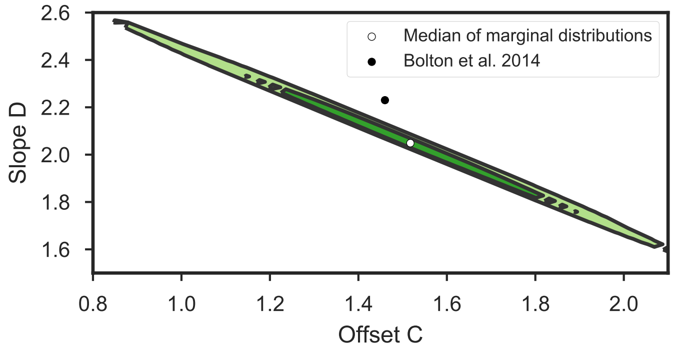


Figure 2.18. Comparison of the b_0-T_0 calibration values with Bolton et al. (2014) at $z = 2.4$ in terms of the joint distribution of C, D. The 68% confidence levels are plotted in dark green and 95% in light green. Figure credit: Hiss et al. (2018).

We tested if the evolution of T_0/γ is consistent with a peak/dip by comparing χ^2 distributions $P(\chi^2|\mathrm{dof})$ of fits to our measurements, where dof is the number of degrees of freedom. For this purpose, we use a four-parameter piecewise linear function $f(z)$ of the form

$$f(z) = \begin{cases} s_1(z - z_{\rm br}) + o & z < z_{\rm br} \\ s_2(z - z_{\rm br}) + o & z \geq z_{\rm br}, \end{cases} \tag{2.13}$$

shown in light gray in Figure 2.20, that describes two linear functions parameterized with two slopes s_1 and s_2, an offset o and a break redshift $z_{\rm br}$. For comparison, we also compute the best fits for a two-parameter linear evolution and a constant. For the evolution of T_0, a piecewise linear function with a best-fit break at $z_{\rm br} = 2.9$ results in a $P(\chi^2|\mathrm{dof}) = 0.097$ for 4 dof. The best-fit linear evolution results in $P(\chi^2|\mathrm{dof}) = 6.5 \times 10^{-4}$ for 6 dof, while no evolution in T_0 results in $P(\chi^2|\mathrm{dof}) = 2.4 \times 10^{-4}$ for 7 dof. This provides some indication that our measurements prefer a model with a peak in the temperature. In the case of γ, a piecewise linear function with a break at $z_{\rm br} = 3.0$ results in a $P(\chi^2|\mathrm{dof}) = 0.12$. This is only slightly better than the $P(\chi^2|\mathrm{dof}) = 0.06$ that we observe for the linear evolution model and best-fit constant $\gamma = 1.4$ with

Table 2.4. Goodness of fit for different models.

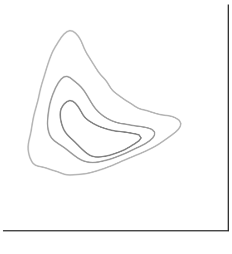
Function	Param.	dof	$P(\chi^2\vert\mathrm{dof})$	$\chi^2_{\mathrm{red.}}$
Constant	T_0	7	2.4×10^{-4}	3.67
	γ	7	0.01	2.13
Linear	T_0	6	6.5×10^{-4}	3.56
	γ	6	0.06	1.42
Piecewise Linear	T_0	4	0.097	1.30
	γ	4	0.12	1.11

[a] Fit for different model types (first column) to the evolution of the parameters of the temperature-density relation (second column) measured in this chapter. The goodness of the fit is expressed as the value of the χ^2 distribution given the number of degrees of freedom (dof, third column), $P(\chi^2|\mathrm{dof})$ (fourth column). Additionally, we show the reduced χ^2 (fifth column).

$P(\chi^2|\mathrm{dof}) = 0.01$. This suggests that a dip in the evolution of γ is slightly preferred given the size of our error bars. A comparison of all fits, including the reduced χ^2, is given in Table 2.4.

The peak in T_0 is suggestive of a late-time $z \sim 3$ process heating the IGM. The reionization of singly ionized helium He II (He II → He III) by a QSO-driven metagalactic ionizing background is the most obvious candidate that would produce such an effect. It has also been argued that He II reionization ends around $z \sim 3$ (Worseck et al. 2011, 2018), which coincides with the redshift at which our measurements of T_0 appear to peak (Upton Sanderbeck et al. 2016; Puchwein et al. 2015; Oñorbe et al. 2017).

Additionally, if the temperature increase comes about independently of the density of the IGM, i.e. the photoionization rate is much higher than the recombination rate everywhere, then the IGM is driven to a temperature-density relation that is close to isothermal (see nonequilibrium simulations in Puchwein et al. 2015). This causes a flattening of the temperature-density relation, which corresponds to a dip in the evolution of γ. In case that the amount of heating is proportional to the neutral fraction of the gas, e.g. high-density regions with higher recombination rates experience more



heating, then the flattening of γ is expected to be less prominent (Puchwein et al. 2015). Given that our data only slightly prefer a dip in γ over a constant evolution, we cannot clearly disentangle these scenarios. Furthermore, the evolution of γ seems to be consistent with a constant if we apply an LD minimization method for the cutoff fitting (see § 2.4.4).

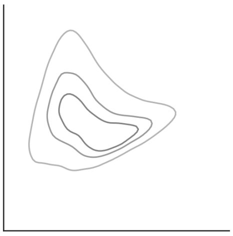
After He II reionization and its concomitant heat injection are complete, the IGM is expected to cool down on a timescale of several hundred Myr (Hui & Gnedin 1997; McQuinn & Upton Sanderbeck 2016), or $\Delta z \sim 1.0$, and asymptote to a T_0 and γ set by the interplay of the photoionization heating and adiabatic cooling, independent of the details of reionization. Due to this process, the IGM is heated by photoionization and then left to cool by cosmic expansion once most of the He II is ionized. This physical picture is consistent with our measured evolution of T_0 and γ.

2.4.2. Comparison with Models

In Figure 2.20, we compare our measurements to a semi-analytical model by Upton Sanderbeck et al. (2016) constructed by following the photoheating history of primordial gas (red solid line) and nonequilibrium reionization simulations by Puchwein et al. (2015). We also compare to different thermal histories from the THERMAL suite (blue curves from Nyx simulations; Almgren et al. 2013; Lukić et al. 2015). Each Nyx simulation was run using different UVB and applying different heat inputs to create three different thermal histories following the method introduced in Oñorbe et al. (2017): (1) no He II reionization (blue solid line), (2) He II reionization ending at $z = 3$ with a temperature input $\Delta T_{\mathrm{HeII}} = 3 \times 10^4$K (blue dashed line), and (3) He II reionization ending at $z = 5.5$ with a temperature input $\Delta T_{\mathrm{HeII}} = 1.5 \times 10^4$K (blue dot-dashed line).

First, we note that if He II reionization never happened or ended at high redshift, then the simulations suggest that T_0 would be ~ 10000 K lower than our measurements at $z = 3$. Furthermore, in agreement with the models, the temperature at mean density decreases at $z < 3$. Our measurements suggest that T_0 is higher than the Upton Sanderbeck et al. (2016) fiducial model and Puchwein et al. (2015) nonequilibrium simulation, with the difference that the nonequilibrium simulation peaks at higher redshift.

The evolution of γ from Upton Sanderbeck et al. (2016) shows a dip at $z = 3$ at nearly the same position as our lowest measurement. The γ dip in the nonequilibrium simulation appears at higher redshifts, coinciding with the corresponding peak in T_0. The thermal evolution of the Nyx simulation (2), with He II reionization at $z = 3$, shows a larger γ at this redshift because the heating due to He II reionization in the model is more extended and already started at higher redshift (see Oñorbe et al. 2017 for more details on the models and their intrinsic limitations). In summary, our measurements of T_0 are suggestive of a heating event taking place between $z = 3.4$ and 3.



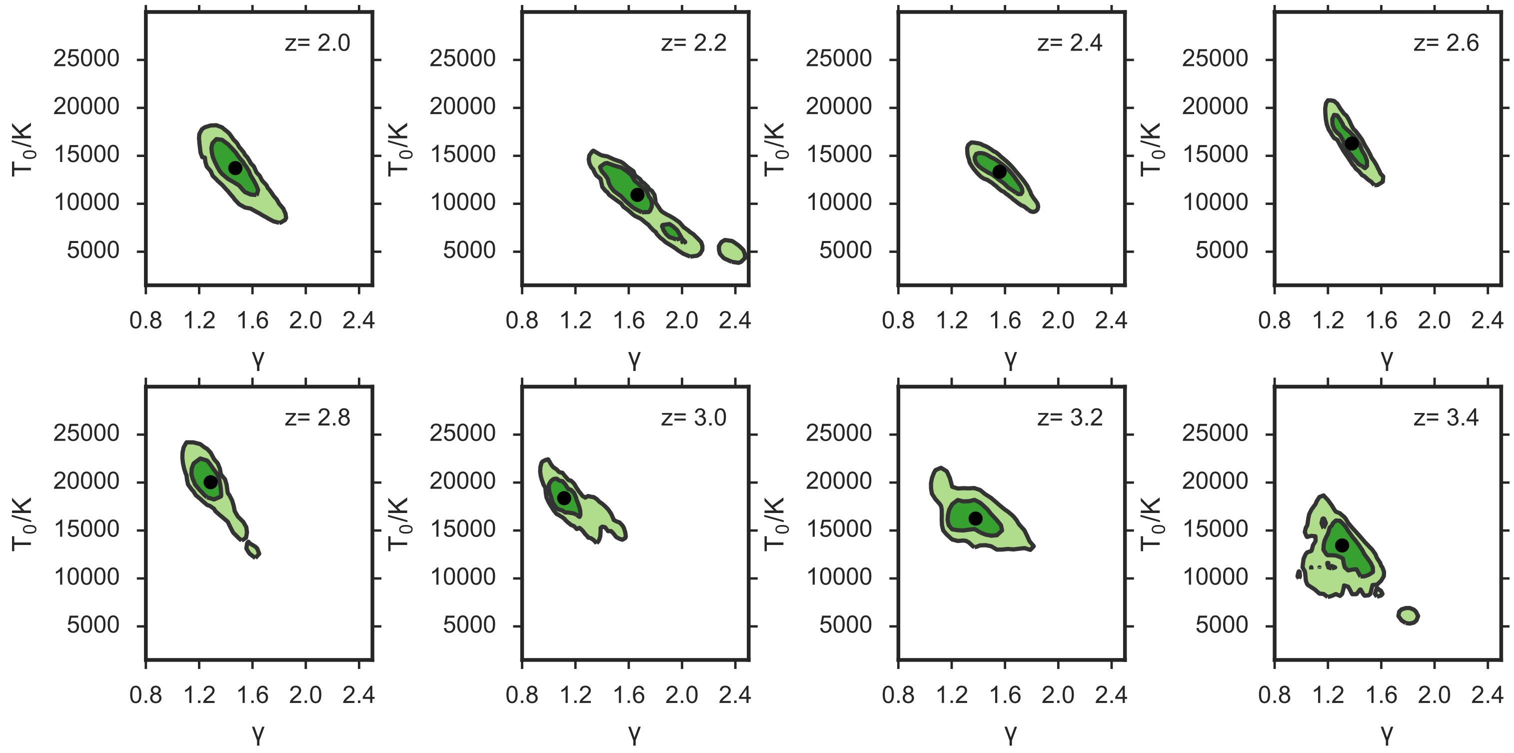


Figure 2.19. Resulting $p(T_0, \gamma)$ PDFs. This is the combination of our data-measured $p\,(b_0, \Gamma)$ PDFs with the simulation-extracted calibration $p(\kappa)$ and $p(A, B)$ PDFs. Each panel represents a redshift bin of size $\delta z = 0.2$. The 68% confidence levels are plotted in dark green and 95% in light green. The black points correspond to the median of the marginal distributions of T_0 and γ.

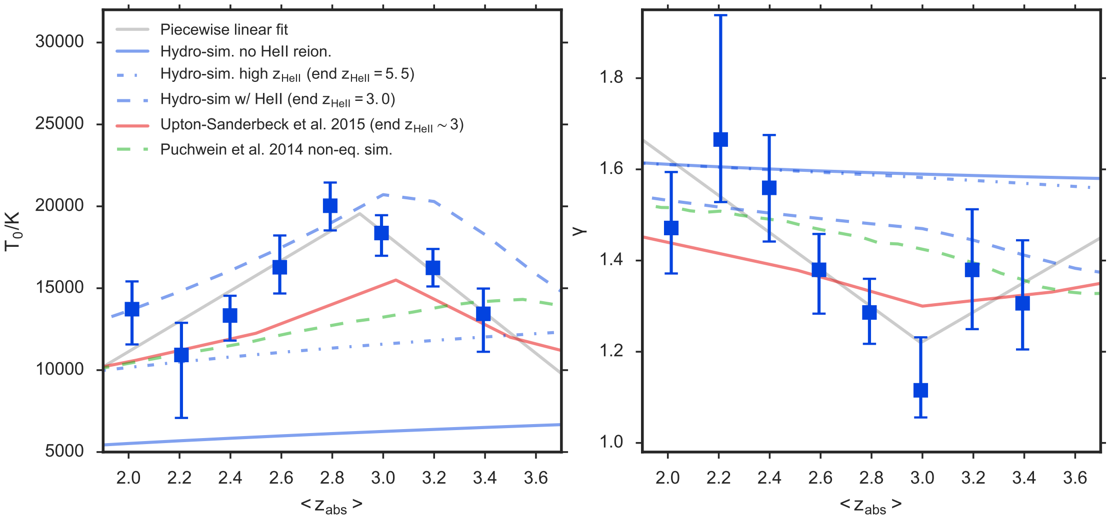
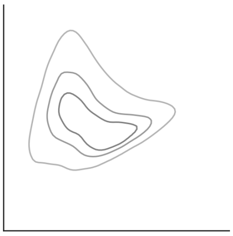


Figure 2.20. Evolution of $\gamma(z)$ and $T_0(z)$ compared to models. The measurements come from the marginal distributions of $p(T_0, \gamma)$ generated by combining all points in the bootstrapped PDF $p\,(b_0, \Gamma)$ from the data cutoff fits with all points in the calibration PDFs $p(C, D)$ and $p(\kappa)$ using eqns. 2.10 and 2.11. The error bars are estimated using the 16th and 84th percentiles of the marginal distributions of $p(T_0, \gamma)$. For comparison, we plot three different Nyx simulations from Oñorbe et al. (2017), a semi-analytical model by Upton Sanderbeck et al. (2016), and a nonequilibrium reionization simulation by Puchwein et al. (2015). A best-fit four-parameter piecewise linear function (described in § 2.4.1) is shown in light gray. Figure credit: Hiss et al. (2018).

2.4.3. Comparison with Previous Work

We can directly compare our cutoff fitting results at $z = 2.4$ with those presented in Rudie et al. (2012a), shown in the $z = 2.4$ panel of Figure 2.7. At $z = 2.4$, our bootstrapped cutoff position measurement yields $\Gamma = 1.17 \pm 0.03$, which is in good agreement with $\Gamma = 1.156 \pm 0.032$ measured by Rudie et al. (2012a). If we evaluate their measurement $b_{0R} = b(N_{\rm HI,0} = 10^{13.6}\ {\rm cm}^{-2}) = 17.56 \pm 0.4\ {\rm km\,s}^{-1}$ at the position of our $N_{\rm HI,0}(z = 2.4) = 10^{13.22}\ {\rm cm}^{-2}$ while keeping their Γ fixed, this measurement becomes $b'_{0R} = 15.32 \pm 0.55\ {\rm km\,s}^{-1}$. Our measurement $p\,(b_0, \Gamma)$ marginalized over Γ (with $b_0 = 18.68^{+0.74}_{-1.07}\ {\rm km\,s}^{-1}$) is more than 3σ higher than this value, indicating tension between our measurements and those of Rudie et al. (2012a) in terms of b_0. This discrepancy is probably due to a different implementation of the cutoff and Voigt profile fitting algorithms used. We performed a cutoff fit our data at $z = 2.4$ using an LD minimization algorithm and although it tends to lead to smaller values of b_0, we



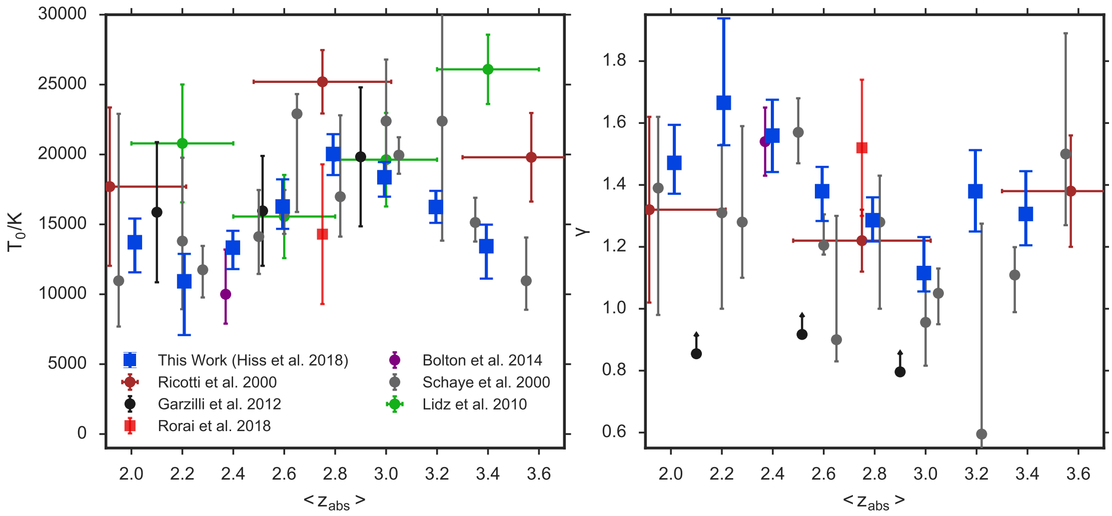


Figure 2.21. Evolution of $\gamma(z)$ and $T_0(z)$, based on the marginal distributions of the $p(T_0, \gamma)$ PDFs, compared to previous measurements. Figure credit: Hiss et al. (2018).

cannot reproduce this low cutoff.

The left panel of Figure 2.21 shows a comparison of our T_0 evolution with previous measurements. Our measurements of T_0 are in good agreement with those of Schaye et al. (2000). We disagree with Ricotti et al. (2000) at $z > 2.4$, where we tend to measure significantly lower temperatures.

Note that our T_0 measurement agrees with that of Bolton et al. (2014), who recalibrated the cutoff measurement of Rudie et al. (2012a) at $z = 2.4$. The fact that we measure inconsistent values of b_0 should lead to inconsistent values in T_0. However, given the difference in our calibration values D, C, this inconsistency is alleviated. Furthermore, Bolton et al. (2014) added a systematic error contribution to their statistical uncertainty in T_0 due to scatter in the $N_{\rm HI}$-overdensity relation in their simulations that led to a 0.2 dex uncertainty in $N_{\rm H\,I,0}$. When adopting values of $N_{\rm H\,I,0}$ that are 0.2 dex above/below the values determined in § 2.3.2 self-consistently in our simulations and data, we observe that the calibration compensates for the choice of $N_{\rm H\,I,0}$, leading to negligible changes in the final results. This is explicitly tested in the appendix A.5. In other words, choosing a higher value of $N_{\rm H\,I,0}$ will increase the value of b_0 almost equally in the data and simulations. Note that this is only true as long as the γ-dependency in eqn. 2.10 remains small. Since our uncertainty in T_0 is dominated by the statistical error of b_0, we adopt no systematic uncertainty term for $N_{\rm H\,I,0}$.

Our measurements are in good agreement with the wavelet amplitude PDF measurements by Garzilli et al. (2012). Comparison with wavelet decomposition measurements by Lidz et al. (2010) in our redshift range shows agreement at intermediate redshifts but $> 2\sigma$ disagreement at $z \sim 2.2$ and 3.4. An analogous disagreement was observed previously in Becker et al. 2011 (in the context of curvature measurements), but its source remains unclear.

We show a comparison of our γ values with other measurements in the literature in the right panel of Figure 2.21. Our measurements of γ agree with those of Schaye et al. (2000) and Ricotti et al. (2000). We also observe a low values of γ at redshifts around $z = 3$.

Our measurement of γ at $z \simeq 2.4$ agrees with that of Bolton et al. (2014). This was expected, given the agreement with Rudie et al. (2012a) in terms of Γ.

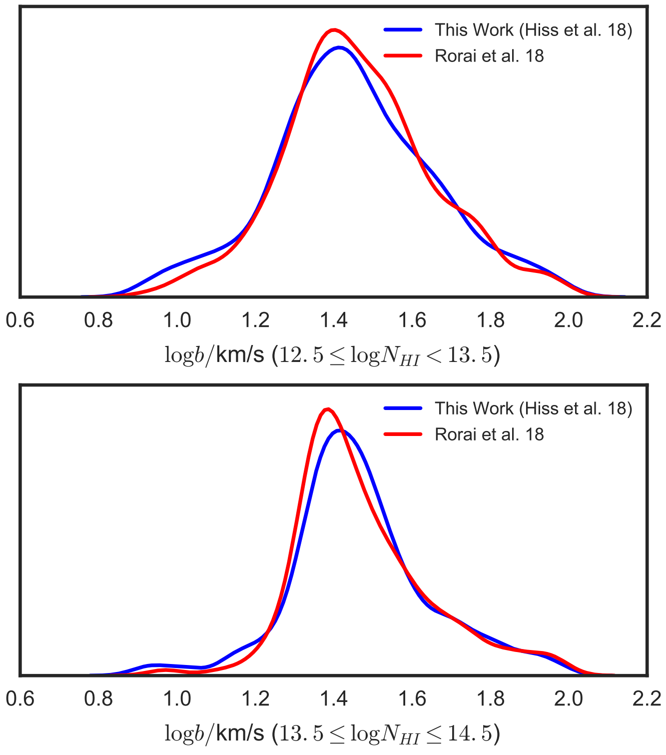


Figure 2.22. Comparison of normalized $\log b$ distributions of data in Rorai et al. (2018) in the redshift bin $2.55 \leq z \leq 2.95$ and this work in the redshift bin $2.7 \leq z < 2.9$. Figure credit: Hiss et al. (2018).

2.4.4. Detailed Comparison with Rorai et al. (2018)

Recently, a study by Rorai et al. (2018) reported measurements of the thermal state of the IGM in the redshift interval $2.55 \leq z \leq 2.95$ that resulted in values of T_0 and γ

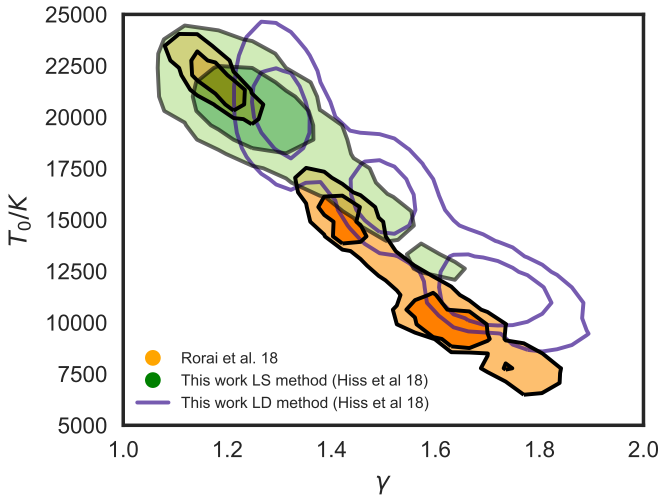


Figure 2.23. Comparison of T_0 and γ contours in Rorai et al. (2018) and this work (LS method in green, LD method in purple) at $z = 2.8$. The contours correspond to the 68% and 95% confidence regions. Figure credit: Hiss et al. (2018).

that are only marginally consistent with our measurement at $2.7 \leq z < 2.9$. To test if the source of this discrepancy originates from the way in which the Voigt profile algorithm was applied to the respective datasets, we plotted the line-width distributions for both our line lists for two intervals of 1 dex in $N_{\rm HI}$ within the cutoff fitting range. The distributions shown in Figure 2.22 are essentially identical. Thus, any difference in the resulting thermal parameters must originate in the cutoff fitting procedure due to contamination, spurious lines, or differences in the calibration.

In Figure 2.23 a direct comparison of the contours of T_0 and γ shows that Rorai et al. (2018) measured a multimodal joint distribution $p(T_0, \gamma)$ (orange) while our measurement (green) recovers only the peak with the highest T_0 and lowest γ. The main difference between the two methods is that we perform a LS minimization fit at each iteration of the cutoff fitting procedure, while Rorai et al. (2018) performs an LD minimization fit. Our algorithm tends to converge to the peak corresponding to high b_0 and low Γ, resulting in this difference.

For comparison we rerun our measurements, this time applying an LD minimization fit for both our data and simulations. Due to unstable behavior of the LD minimization method at some redshifts, we applied no 2σ outlier rejection (§ 2.1.5) to our data b-$N_{\rm HI}$ distribution when applying this method. We show the resulting $p(T_0, \gamma)$ contours at $z = 2.8$ in purple in Figure 2.23. The results of the evolution of T_0 and γ are shown in Figure 2.24. Essentially, the main difference between the two methods when applied to our data is that we see extended uncertainties at $z = 2.6$ and $z = 2.8$ that originate from multimodal distributions $p(T_0, \gamma)$. Furthermore, the redshift evolution of γ is consistent with a constant $\gamma \simeq 1.4$.

As in Rorai et al. (2018), when using the LD minimization method, we observe a multimodal $p(T_0, \gamma)$ distribution at $z = 2.8$ (also at $z = 2.6$) in the data that results from a multimodal $p\,(b_0, \Gamma)$ measurement. When dealing with simulated b-N_{HI} distributions, both methods lead to unimodal solutions. This opens up the question of whether these multiple peaks in the inference of the cutoff parameters are a real feature due to multimodality in the temperature or an artifact of the cutoff fitting procedure due to unknown systematics in the data.

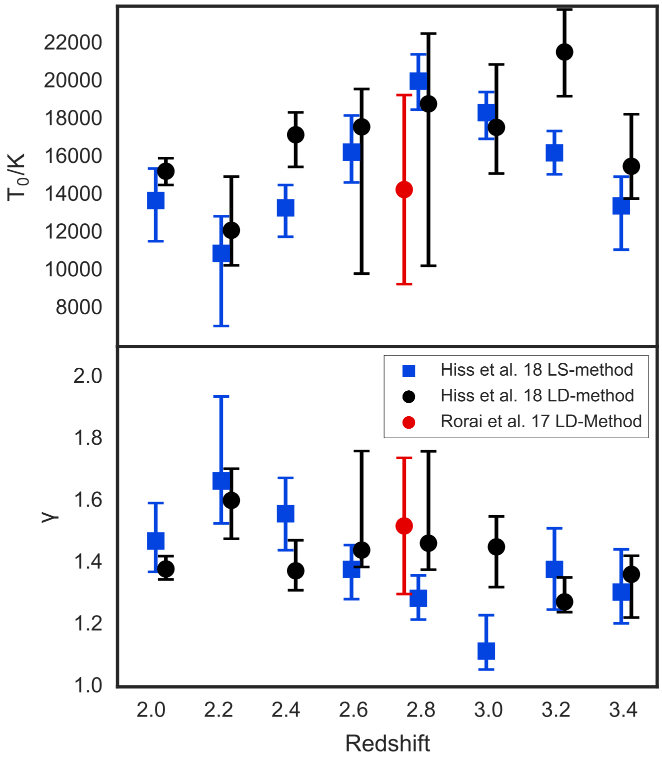
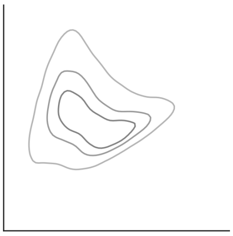


Figure 2.24. Comparison of the marginalized T_0 and γ in Rorai et al. (2018) (red) and this work (blue). We also ran our procedure using an LD minimization cutoff fitting procedure (black). The main difference between the methods is that the LS minimization method used in this chapter does not show a multimodal structure at $z = 2.6$ and $z = 2.8$. Also, the evolution of γ is consistent with a constant, not showing a dip at $z = 3$. Figure credit: Hiss et al. (2018).



2.4.5. Evolution of the Temperature at Optimal Density

The temperature-density relation is traditionally normalized at mean density. However, at different redshifts, an optical depth of ~ 1 in the Lyα forest traces different overdensities. Based on this, Becker et al. (2011) introduced the mean curvature statistic $\langle|\varkappa|\rangle$, which is a probe of the thermal state of the IGM that is related to the temperature at optimal density $T(\bar{\Delta}) = T(\rho_{\mathrm{opt}}/\rho_0)$ independent of γ.

For a fair comparison of our measurements with those from Becker et al. (2011), we apply another transformation on our measurements so we can look at the evolution of the temperature of the IGM in terms of the temperature at the optimal density $T(\bar{\Delta})$. If we rewrite the temperature-density relation in terms of $T(\bar{\Delta})$,

$$T(\bar{\Delta}) = T_0\bar{\Delta}^{\gamma-1} \tag{2.14}$$

we are able to combine our $p(T_0, \gamma)$ PDF with measurements of $\bar{\Delta}$ by Becker et al. (2011), which have no reported uncertainties. Plugging in all pairs of (T_0, γ) from $p(T_0, \gamma)$ into eqn. 2.14 in combination with a fixed value of $\bar{\Delta}$ (linearly interpolated based on Becker et al. 2011 to match our redshift bins) allows us to generate $p(T(\bar{\Delta}), \gamma)$ PDFs for each redshift. This approach takes into account any covariance with respect to γ in our measurements. The resulting $p(T(\bar{\Delta}), \gamma)$ contours are shown in Figure 2.25. We note that covariance between $T(\bar{\Delta})$ and γ is diminished compared to that between T_0 and γ (see Figure 2.19 for comparison) when taking our measurements to $T(\bar{\Delta})$ space. However, note that our $T(\bar{\Delta})$ contours are correlated with γ in most redshift bins.

Given $p(T(\bar{\Delta}), \gamma)$ joint distributions, we can marginalize out γ and compare $T(\bar{\Delta})$ directly to Becker et al. 2011 and Boera et al. 2014 (also computed using the mean curvature method). Our 68% confidence regions for $T(\bar{\Delta})$ as a function of redshift are shown in Figure 2.26. A comparison with Becker et al. (2011) is not completely straightforward, given that the redshift bin sizes are different, and we are also linearly interpolating their $\bar{\Delta}$ values. Broadly speaking, we see agreement with Becker et al. (2011) and Boera et al. (2014) at 1σ level at $z < 2.4$, 3.0, and 3.4, as well as generally higher temperatures at $2.4 \leq z \leq 3.2$ that disagree at the $> 2\sigma$ level. Given the method dependency (see § 2.4.4) and other systematics associated with cutoff fitting, the difference might not be as significant as it appears, once these effects are properly quantified. Additionally, if one included uncertainties in $\bar{\Delta}$, it would further alleviate this tension. One possible effect that could be playing a role is that the curvature statistic is sensitive to metals in the Lyα forest that do not get masked; i.e. metal contamination leads to lower values of $T(\bar{\Delta})$ (Boera et al. 2014). This effect is potentially more prominent at higher redshifts, where blending of Lyα forest lines makes it more difficult to identify all metals. Our analysis is, in principle, less sensitive to metals given our 2σ rejection procedure adopted before cutoff fitting, but the exact source of this discrepancy remains unclear.

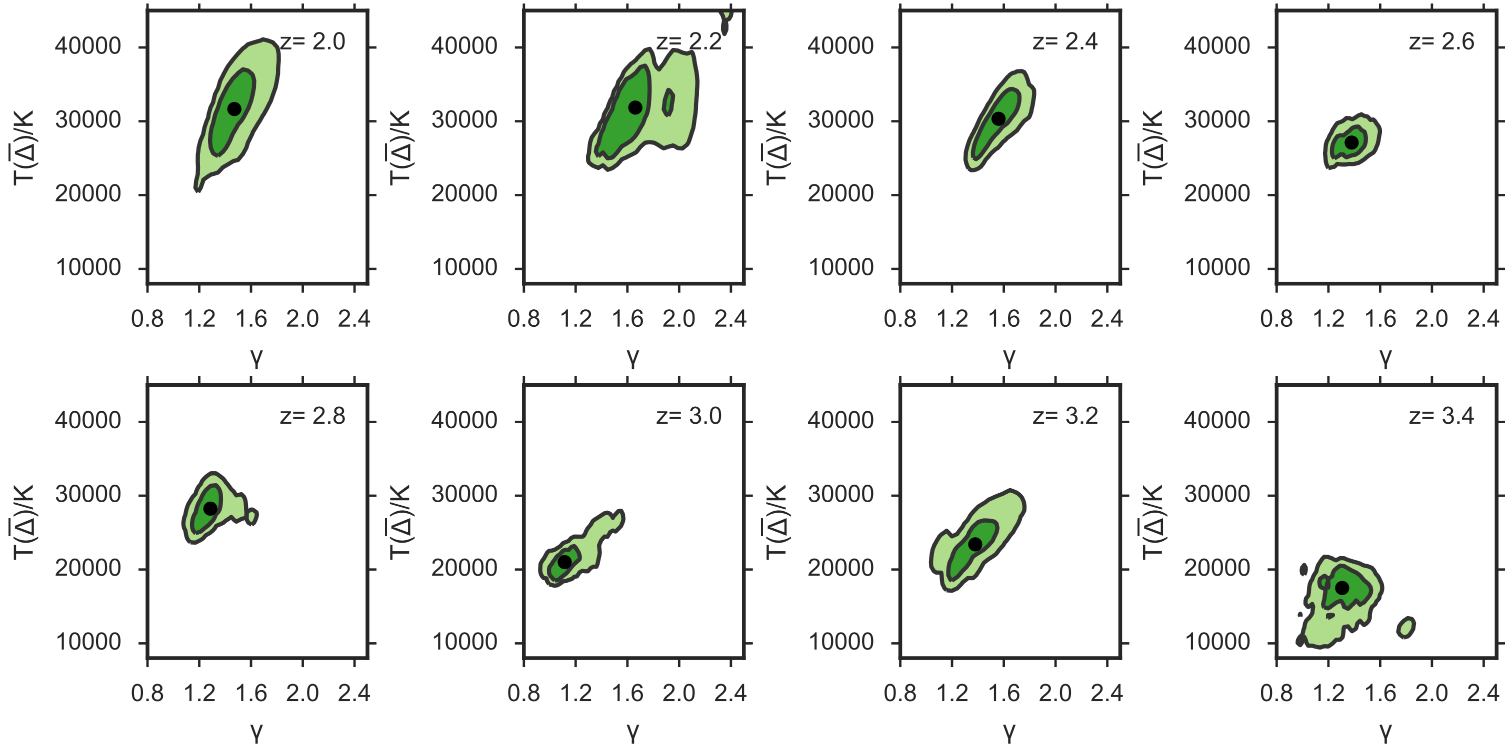
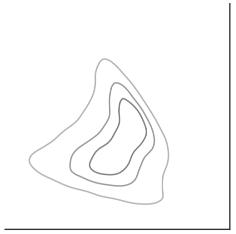


Figure 2.25. Resulting $p(T(\bar{\Delta}), \gamma)$ PDFs. This is the combination of our calibrated $p(T_0, \gamma)$ PDFs at every redshift with the measurements of $\bar{\Delta}$ by Becker et al. (2011). The 68% confidence levels are plotted in dark green and 95% in light green. The black points correspond to the median of the marginal distributions of $T(\bar{\Delta})$ and γ.



An overview of all quantities measured and adopted in this chapter is given in Table 2.5. A subset of the measurements on which the distributions $p\,(b_0, \Gamma)$, $p(D, C)$, $p(\kappa)$ and $p(T_0, \gamma)$ are based is available in machine-readable form for all redshifts presented and can be obtained in the Zenodo repository Hiss et al. (2018)[11].

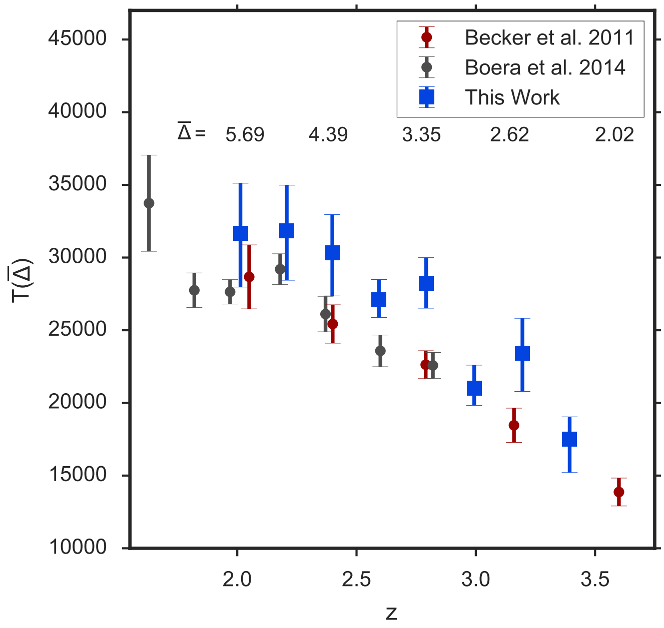


Figure 2.26. Comparison to Becker et al. (2011) and Boera et al. (2014) after combining our $p(T_0, \gamma)$ with the Becker et al. (2011) measurements of $\bar{\Delta}$. We measure a hotter IGM at higher redshifts. Figure credit: Hiss et al. (2018).

---

[11] Url: `https://zenodo.org/record/1285569`

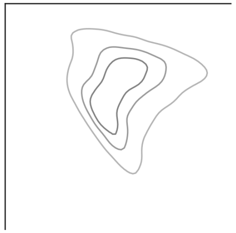


Table 2.5. Measurements and values adopted

z	b_0	D	C	T_0	Γ	κ	γ	$\bar{\Delta}$	$T(\bar{\Delta})$
2.0	$18.22^{+0.97}_{-1.39}$	2.11 ± 0.245	1.48 ± 0.305	13721^{+1694}_{-2152}	$1.14^{+0.03}_{-0.03}$	3.48 ± 0.185	$1.47^{+0.12}_{-0.1}$	5.85	31659^{+3690}_{-3455}
2.2	$16.89^{+1.37}_{-3.12}$	2.1 ± 0.225	1.46 ± 0.285	10927^{+1961}_{-3843}	$1.19^{+0.07}_{-0.04}$	3.6 ± 0.1	$1.67^{+0.27}_{-0.14}$	5.1	31853^{+3415}_{-3126}
2.4	$18.68^{+0.74}_{-1.07}$	2.04 ± 0.19	1.54 ± 0.25	13334^{+1206}_{-1530}	$1.17^{+0.03}_{-0.03}$	3.39 ± 0.11	$1.56^{+0.12}_{-0.12}$	4.4	30335^{+2976}_{-2617}
2.6	$20.41^{+1.02}_{-0.87}$	2.31 ± 0.23	1.19 ± 0.3	16281^{+1940}_{-1601}	$1.12^{+0.02}_{-0.03}$	3.22 ± 0.08	$1.38^{+0.08}_{-0.1}$	3.87	27113^{+1234}_{-1372}
2.8	$22.67^{+0.55}_{-0.6}$	2.38 ± 0.135	1.08 ± 0.175	20036^{+1416}_{-1507}	$1.1^{+0.02}_{-0.02}$	2.95 ± 0.1	$1.29^{+0.07}_{-0.07}$	3.35	28245^{+1729}_{-1750}
3.0	$22.24^{+0.33}_{-0.73}$	2.35 ± 0.19	1.1 ± 0.25	18371^{+1087}_{-1388}	$1.04^{+0.04}_{-0.02}$	2.96 ± 0.11	$1.12^{+0.12}_{-0.06}$	2.95	21002^{+1171}_{-1596}
3.2	$21.65^{+0.4}_{-0.52}$	2.53 ± 0.23	0.84 ± 0.3	16244^{+1153}_{-1135}	$1.13^{+0.04}_{-0.04}$	2.96 ± 0.145	$1.38^{+0.13}_{-0.13}$	2.57	23410^{+2623}_{-2409}
3.4	$20.8^{+0.71}_{-1.27}$	2.97 ± 0.265	0.22 ± 0.35	13439^{+1542}_{-2318}	$1.11^{+0.05}_{-0.04}$	2.72 ± 0.12	$1.31^{+0.14}_{-0.1}$	2.3	17500^{+2289}_{-1542}

Note. — Summary of all quantities measured/used in this work. The first column shows the center of each redshift bin used. The second column shows the median and percentile based errors of the cut-off fitting parameter b_0. The third and fourth columns show the calibration parameters C, D from eqn (2.10). The fifth column shows the resulting T_0 once the calibration is applied. The sixth column shows the median and percentile based errors of the cut-off fitting parameter Γ. The seventh column shows the calibration parameter κ from eqn (2.11). The eigth column shows the resulting γ once the calibration is applied. The ninth column shows the values of the optimal-density that were linearly interpolated from Becker et al. (2011). The last column shows the values of the temperature at optimal density $T(\bar{\Delta})$ constructed using eqn. (2.14).

2.4.6. Caveats

It should be noted that there are a number of assumptions adopted in this study that we summarize as follows.

We assume that the simulated b-$N_{\mathrm{H\,I}}$ distributions are comparable to the ones extracted from the data, or in other words, that the cutoff fitting algorithm will respond similarly in both cases. This is an especially problematic assumption, because metals have to be rejected from our data, which are, by construction, not present in the simulated mock spectra. Therefore, we observe that the b-$N_{\mathrm{H\,I}}$ distributions from mock spectra generate much more concentrated cutoff fitting bootstraps (see Figures 2.11 and 2.12). This effect increases the errors measured in b_0 and Γ in the data, which dominate the error budget of T_0 and γ. Furthermore, our simulations do not account for effects such as multimodality in the temperature-density relation which could play a role, especially at $z > 2.8$.

Another assumption is that the calibrations for T_0 and γ can be done separately, i.e. $p(D, C, \kappa) = p(D, C)p(\kappa)$. This is not necessarily true, as these parameters could be correlated. As we calculated the calibration values on the same bootstrap samples, any correlation is still preserved. We inspected the distributions of $p(\kappa, C)$ and $p(\kappa, D)$ and did not find significant correlation.

In this work, we utilize a LS minimization fitting algorithm in every iteration of the cutoff fitting process. This is a different approach than in previous works, and our final results are sensitive to the method chosen. This aspect is further discussed in § 2.4.4 in the context of the comparison of our work with the results of Rorai et al. (2018).

As mentioned in Schaye et al. (1999), if the reionization process has large spatial fluctuations and the gas has not settled into one temperature-density relation (see Compostella et al. 2013; McQuinn & Upton Sanderbeck 2016), the measurement of the position of the cutoff will be sensitive to the gas with the lowest temperature. If this is the case, the temperature measurements should be treated as lower limits to the average temperature.

2.5. Summary

In this chapter, we assessed the thermal state of the IGM and its evolution in the redshift range $2.0 \leq z \leq 3.4$ using 75 high SNR and high-resolution Lyα forest spectra from the UVES and HIRES spectrographs. We exploited the fact that absorbers that are primarily broadened due to the thermal state of the gas have the smallest Doppler parameters, which results in a low-b cutoff in the b-$N_{\mathrm{H\,I}}$ distribution. We decomposed the Lyα forest of these spectra into a collection of Voigt profiles and measured the position of this cutoff as a function of redshift. We calibrate this procedure using 26 combinations of thermal parameters at each redshift from the THERMAL suite of hydrodynamic simulations with different values of the IGM pressure smoothing scale. We conduct an end-to-end analysis whereby both data and simulations are treated in a self-consistent way, and uncertainties in both the cutoff fitting and the calibration procedure are propagated into our analysis.

The primary results of this chapter are as follows:

1. We see suggestive evidence for a peak in IGM temperature evolution at $z \simeq 2.9$. The temperature at mean density T_0 increases with decreasing redshift over the range $2.9 < z < 3.4$, peaks around $z \simeq 2.9$, and then again decreases with redshift over the range $2.0 < z < 2.9$.

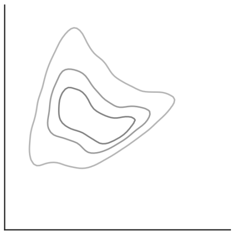
2. When applying our cutoff fitting procedure, the redshift evolution of γ suggests a dip around $z \simeq 3.0$ over a linear or constant evolution model when using a simple piecewise linear evolution model that decreases in the redshift interval $2.9 < z < 3.0$ and increases in the interval $2.0 < z < 3.0$.

3. We observe significantly higher temperatures at mean density $T_0 \simeq 15000 - 20000\,\mathrm{K}$ at $2.4 < z < 3.4$ than the much lower $T_0 \simeq 6000\,\mathrm{K}$ predicted by models for which He II reionization did not take place or compared to the $T_0 \simeq 10000\,\mathrm{K}$ expected if He II reionization ended at very high redshift ($z = 5.5$).

4. In contrast to previous analyses based on the flux PDF (Bolton et al. 2008; Viel et al. 2009), our measurements disfavor negative values of $\gamma - 1$ at high statistical significance. Assuming that the IGM follows a temperature-density relation closely, this means that inverted temperature-density relations are unlikely at $2.0 \leq z \leq 3.4$. Note that the discrepancies with flux PDF measurements can be attributed to an upturn in temperature at low densities and whether the IGM temperature-density relation is multiphased at low densities (Rorai et al. 2017a).

5. Our measurements of T_0 and γ can also be phrased as measurements of $T(\bar{\Delta})$, which is the quantity measured by curvature studies. We find broad agreement with the Becker et al. (2011) and Boera et al. (2014), curvature measurements at $z < 2.4$, 3.0, and 3.4, but we observe higher values of $T(\bar{\Delta})$ in the interval $2.4 \leq z \leq 3.2$.

In summary, both the suggestive peak in the redshift evolution of T_0 at $z \sim 2.9$ and the relatively high IGM temperatures $T \sim 10000 - 20000$ at $2.0 \leq z \leq 3.4$ provide evidence for a process that heated the IGM at $z \sim 3 - 4$. The most likely candidate responsible for this thermal signature is He II reionization.



3. A Novel Method for Fitting the b-$N_{\mathrm{H\,I}}$ distribution

You don't feel or even see it
But I can't help saying, my friend,
That a new change will soon happen.
And what some time ago was new, young,
Today is old,
And we all need to rejuvenate

- Belchior, *Velha Roupa Nova* 1976

The conventional method for measuring thermal parameters using the joint distribution of column densities and Doppler parameters (b-$N_{\mathrm{H\,I}}$ distribution) of absorbers in the Lyα forest in a particular redshift interval relies on the measurement of the thermal state dependent lower cutoff in this distribution (see Schaye et al. 1999; Ricotti et al. 2000; McDonald et al. 2001; Schaye et al. 2000; Rudie et al. 2012a; Bolton et al. 2014; Garzilli et al. 2015; Rorai et al. 2018; Hiss et al. 2018; Telikova et al. 2018; Garzilli et al. 2018), set primarily by the minimal broadening associated with the temperature of the absorbers. We explored this method extensively in Chapter 2.

Although it constitutes a powerful tool for measuring the thermal state of the gas, the cutoff fitting technique has a series of inherent disadvantages, the main one being that the position of the cutoff is fitted using an iterative technique which excludes absorbers from the distribution. This means that a small number of absorbers is effectively used for measuring the position of the cutoff. This results in a diminished sensitivity of the method on the total number of absorbers in the dataset, once the distribution is well populated (Schaye et al. 1999). In addition, narrow metal line absorbers, which are difficult to completely identify and mask, can result in significant contamination around the cutoff. These features can compromise the precision with which the cutoff can be determined, and add systematic uncertainties which are difficult to control. Another complication of this method, as shown in § 2.4.4 in the context of the comparison with the results by Rorai et al. (2018), is that the choice of cutoff fitting method (i.e. least-squares or mean-deviation minimization) can lead to significantly different T_0 and γ measurements. All of these problems call for a new method for interpreting the

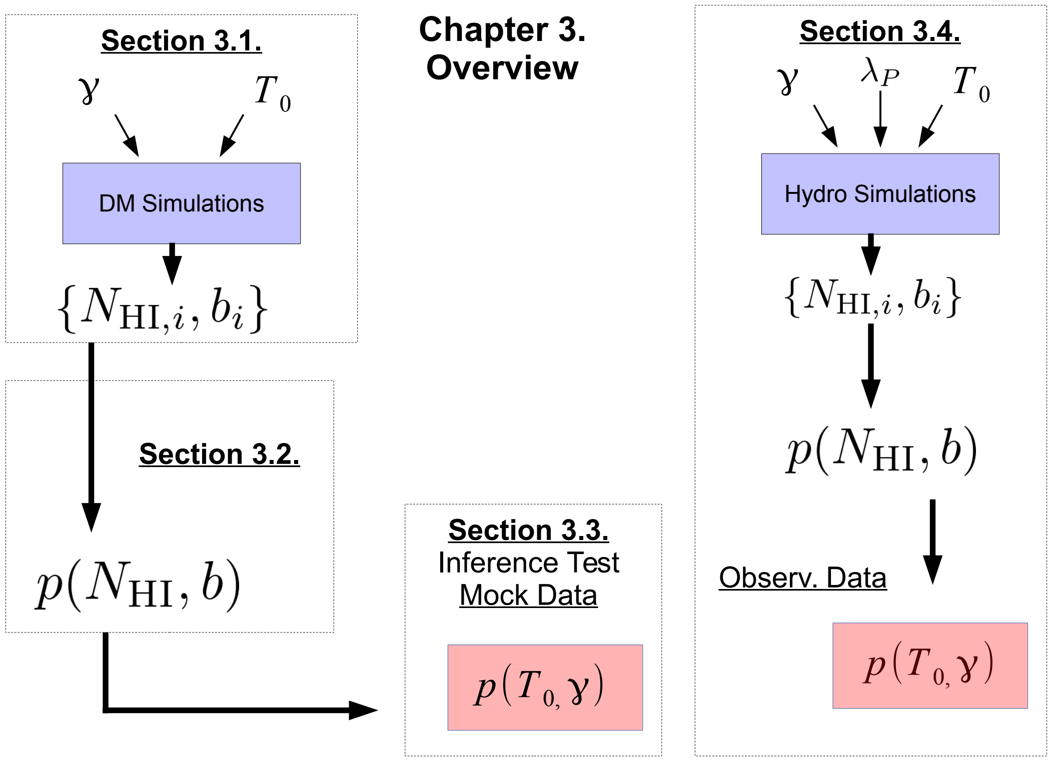


Figure 3.1. Overview of this chapter.

.

information about the thermal state of the IGM encoded in the b-N_{HI} distribution.

In this chapter we introduce, test, and apply a new method for constraining T_0 and γ using the b-N_{HI} distribution. The main difference with the traditional cutoff fitting approach is that we model the entire distribution, and thus bypass the complications associated with quantifying the position of a lower cutoff. While other studies employed a parametric description of the full b-N_{HI} distribution in order to carry out measurements of the parameters of the TDR (see e.g. Ricotti et al. 2000; Telikova et al. 2018), we instead construct smooth probability density functions (PDF) of simulated b-N_{HI} distributions using a non-parametric approach. These PDFs can then be used as models for conducting inference. Here we should note that all results presented in our proof of concept concern T_0 and γ alone and do not marginalize over other parameters. The scope of this study is to demonstrate the capabilities of this new approach.

This chapter is divided into two parts (see Figure 3.1 [1]). All the tests in the first part

---

[1] For readability, we avoided log quantities in this diagram as well as in its description in this introduction. In actuality, we will work with $\log N_{\mathrm{HI}}$, $\log b$ and $\log T_0$

are carried out using DM-only simulations for convenience, while the second part uses hydrodynamic simulations. This chapter's structure is as follows:

1. We introduce our DM-only simulations with different input T_0 and γ values, as well as the mock data generation procedure in § 3.1. From these mock data we measure b-$N_{\rm HI}$ distributions $\{N_{\rm HI,\it i}, b_i\}$ by applying Voigt profile fitting. Our new method for constructing a smooth model of the $P(\log N_{\rm HI}, \log b)$ based on the discrete $\{N_{\rm HI,\it i}, b_i\}$ using kernel density estimation is described in § 3.2. We also discuss how it can be used to carry out inference of $p(T_0, \gamma)$. In § 3.3, we carry out an inference test by measuring $p(T_0, \gamma)$ from many mock data realizations to explore the robustness of this technique.

2. We apply the workflow described in the first part to observational data at $z = 2$ in § 3.4, with the difference that THERMAL hydrodynamic simulations are used for generating model b-$N_{\rm HI}$ distributions.

We discuss and summarize our results in § 3.5.

3.1. Simulations

In this section we describe how we generate simulated Lyα forest spectra with different combinations of the underlying thermal parameters that govern the IGM. Specifically, we wish to generate a grid of T_0, γ at a fixed λ_P to understand how the corresponding shape of the b-$N_{\rm HI}$ distribution changes as a function of the thermal parameters T_0, γ, i.e. $P(\log N_{\rm HI}\ , \log b \mid \log T_0, \gamma)$. Certainly the choice of λ_P has an effect on the shape of the b-$N_{\rm HI}$ distribution, as shown in Garzilli et al. (2015), meaning that one should consider $P(\log N_{\rm HI}, \log b | T_0, \gamma, \lambda_P)$. For the sake of simplifying the analysis for an initial proof of concept, we will test our method at a fixed λ_P. Note that all cosmological length scales in this chapter are given in comoving units.

For generating our T_0, γ grid, we create mock spectra using a snapshot of a DM-only simulation at $z = 2$. Although it is well known that spectra based on approximations to a full hydrodynamic simulation are limited in their ability to accurately represent the IGM (Gnedin & Hui 1998; Meiksin & White 2001; Viel et al. 2006; Sorini et al. 2016), we opt to use DM only simulations first, as they allow us to run large numbers of thermal models in a computationally feasible time, allowing us to generate dense thermal grids. This approach should suffice for initial tests, as both mock data and models are generated from the same sort of simulation and we are mainly interested in generating a method that is sensitive to thermal-state-dependent changes in the shape of the b-$N_{\rm HI}$ distribution. We expand our analysis with the use of hydrodynamic simulations in § 3.4, which is a necessary step when dealing with actual observational data.

Our simulation provides the DM density and velocity fields calculated using an updated version of the TreePM code from White et al. (2002) that evolves $N_p = 2048^3$ collisionless, equal-mass particles ($M_p = 2.5 \times 10^5 M_\odot$) in a periodic cube of side length $L_{\mathrm{box}} = 30\,\mathrm{Mpc}h^{-1}$ with a Plummer equivalent smoothing of $1.2\,\mathrm{kpc}h^{-1}$ (similar to Rorai et al. 2013). The cosmology used in the simulations is consistent within 1σ with the 2013 Planck release (Planck Collaboration et al. 2014) with $\Omega_\Lambda = 0.691$, $\Omega_m = 0.309$, $\sigma_8 = 0.829$, $\Omega_b h^2 = 0.022$, $n_s = 0.961$, and $h = 0.678$.

In order to model lines of sight through the IGM, we extract skewers from our simulation that run parallel to one of the box axes and apply the recipe described below. A pseudo-baryonic field is generated by smoothing the DM density and velocity fields. This smoothing mimics the effect of Jeans pressure smoothing of the gas, i.e. accounts for the fact that small-scale structure is suppressed in the baryonic matter distribution owing to finite gas pressure (Gnedin & Hui 1996, 1998; Kulkarni et al. 2015). We choose to smooth the DM field with a constant (instantaneous) filtering scale λ_P. This is done by convolving the density and velocity fields in real space with a cubic spline kernel of the form

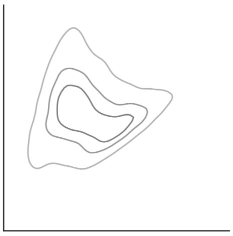
$$K(r, R_P) = \frac{8}{\pi R_P^3} \begin{cases} 1 - 6\left(\frac{r}{R_P}\right)^2 + 6\left(\frac{r}{R_P}\right)^3 & \frac{r}{R_P} \leq \frac{1}{2} \\ 2\left(1 - \frac{r}{R_P}\right)^3 & \frac{1}{2} < \frac{r}{R_P} \leq 1 \\ 0 & \frac{r}{R_P} > 1 \end{cases} \tag{3.1}$$

with a smoothing parameter R_P. This function closely resembles a Gaussian with $\sigma \sim R_P/3.25$ in the central regions, which defines our pressure smoothing scale $\lambda_P = R_P/3.25$. Given the characteristics of our simulations, the mean interparticle separation $\Delta\ell = L_{\mathrm{box}}/N_p^{1/3}$ allows us to resolve values of $\lambda_P \gtrsim 20\,\mathrm{kpc}$ (Rorai et al. 2013). For all DM-only-related models used in this chapter, we will adopt a fixed value of $\lambda_P = 73.3\,\mathrm{kpc}$, which is consistent with the measurement by Rorai et al. (2013) at $z = 2$.

Under the assumption that the IGM is highly ionized and in photoionization equilibrium, we can construct an Lyα optical depth field in real space based on the smoothed DM density field using the fluctuating Gunn-Peterson approximation (FGPA Weinberg et al. 1997; Croft et al. 1998)

$$\tau(x) \propto n_{\mathrm{HI}}(x) \propto T_0^{-0.7} \rho(x)^{2-0.7(\gamma-1)}, \tag{3.2}$$

where x is the particle position in real space. In order to account for the effects of thermal broadening and peculiar velocities of the gas on the optical depth, we compute the redshift-space optical depth by convolving the real-space optical depth with a Gaussian profile. This is an approximation to the actual Voigt profile and is characterized by a thermal width $b = \sqrt{2k_B T/m_{\mathrm{HI}}}$, (where m_{HI} is the hydrogen atom mass, k_b the Boltzmann constant, and T the temperature) and a shift from its real-space position by the longitudinal component of the peculiar velocity. This way we can choose T_0 and



γ, i.e. impose a deterministic power-law TDR onto the simulation. This allows us to generate mock spectra with different sets of underlying thermal parameters T_0 and γ.

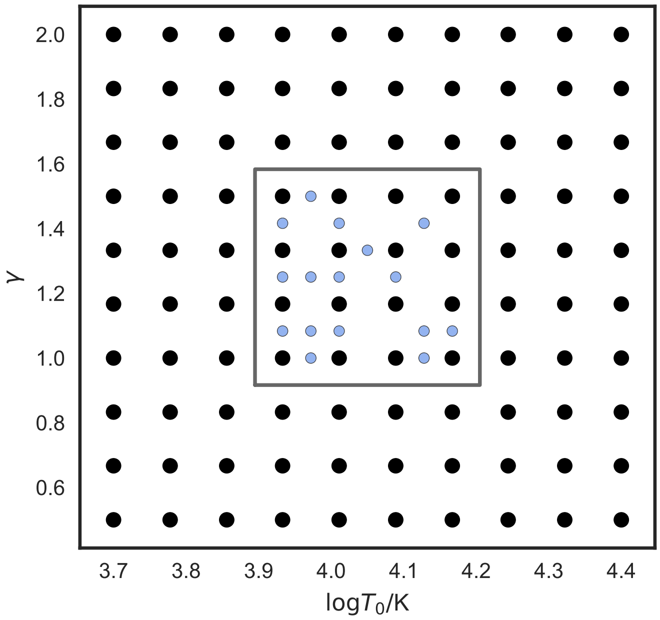


Figure 3.2. Grid of thermal parameters applied to a DM-only simulation at $z = 2$ used to construct a model of the b-$N_{\mathrm{H\,I}}$ distribution. The black points show the combinations of $\log T_0$ and γ imposed on our simulation ("standard grid"). The square marks the area that will be used for inference tests. The blue points indicate where further models were generated for testing the robustness of the method presented in this chapter ("test grid"). Figure credit: Hiss et al. (2019).

The corresponding flux skewer F, the transmission spectrum along the line of sight, is calculated from the optical depth τ using $F = \exp(-A_r\,\tau)$. Here we introduce a scaling factor A_r that allows us to match our lines of sight to observed mean flux values $\bar{F}$. The mean flux normalization is computed for the full snapshot, i.e. the factor A_r is iteratively changed until the mean flux of the snapshot converges to a desired (measured) mean flux. We apply that value of A_r to all the spectra when generating skewers, so there is one mean flux normalization of the whole box and sight-line-to-sight-line variations are still present in our models. This rescaling of the optical depth accounts for our lack of knowledge of the precise value of the metagalactic ionizing background photoionization rate, and it is done simply to generate more realistic skewers. To this end we choose A_r so that we agree with the effective opacity $\tau_{\mathrm{eff}} = -\ln(\bar{F})$ at $z = 2$ from Faucher-Giguère et al. (2008b), namely, $\tau_{\mathrm{eff}} = 0.127$.

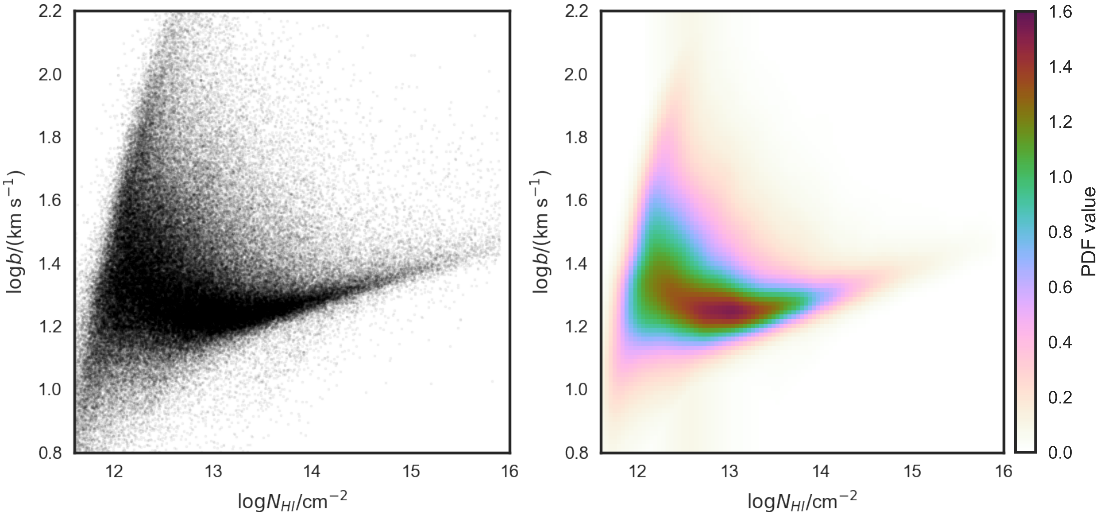
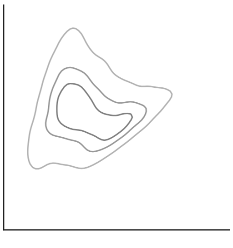


Figure 3.3. **Left**: A b-N_{HI} distribution illustrated as a cloud of points generated by concatenating the `VPFIT` output for 6000 skewers from a DM-only simulation snapshot at $z = 2$ with thermal parameters $(\log T_0, \gamma) = (4.011, 1.333)$. This distribution consists of $\sim 1.5 \times 10^5$ absorbers. **Right**: The KDE-based PDF of the same distribution (as described in § 3.2.1).

3.1.1. Thermal Parameter Grid

Using our simulation snapshot at $z = 2$, we generated 6000 skewers for each of 100 combinations of thermal parameters $\log T_0$ and γ at a fixed $\lambda_P = 73.3\,\mathrm{kpc}$. Figure 3.2 shows the distribution of thermal parameters chosen (black points). We chose to model the thermal parameters on a 10×10 regular grid covering the range $3.7 \leq \log(T_0/\mathrm{K}) \leq 4.4$ and $0.5 \leq \gamma \leq 2.0$, which is dense enough to sample typical uncertainties in T_0 and γ. The number of skewers at each grid point was chosen so that we have enough absorbers to ensure that our estimation of the shape of the b-N_{HI} distribution is converged. This is important, as we will use the absorbers in the b-N_{HI} distribution to estimate $P(\log N_{\mathrm{HI}}, \log b \mid \log T_0, \gamma)$ which we will introduce in § 3.2.1. In this chapter we will refer to this grid as the “standard grid.”

In addition, we generated 16 models between the grid points in the central region of our grid (region marked with the square and blue points in Figure 3.2). These were randomly chosen from a regular grid twice as fine as the standard grid, excluding the points that coincide with it. These additional models will be used in § 3.2.3 to test the robustness of our procedure for generating model b-N_{HI} distributions, as well as our statistical inference (see § 3.3.2). We will refer to these extra models as the “test grid.”



3.1.2. Forward Modeling Noise and Resolution

The technique presented in the following section is based on the sensitivity of the shape of the b-$N_{\mathrm{H\,I}}$ distribution on the thermal state of the IGM. Therefore, it is important that instrumental effects that can also affect the shape of the b-$N_{\mathrm{H\,I}}$ distribution, such as noise and spectroscopic resolution, are properly included in the models we wish to compare to data.

To mimic instrumental resolution, we convolve the skewers with a Gaussian with FWHM = $6\,\mathrm{km\,s^{-1}}$, which is the typical resolution delivered by echelle spectrometers (see e.g. the HIRES and UVES dataset presented in § 2.1). Further, we add Gaussian random noise to the skewers assuming a fixed SNR of 63 per resolution element for the purpose of choosing a value comparable to the SNR of the dataset from § 2.1 at $z = 2$.

We apply the exact same Voigt profile fitting scheme described in § 2.1.2 to the 6000 forward-modeled simulated skewers generated for 100 different combinations of T_0, γ. To summarize, Voigt profiles were fitted to our simulated data using `VPFIT` version 10.2[2] (Carswell & Webb 2014). We wrote a fully automated set of wrapper routines that prepare the spectra for the fitting procedure and control `VPFIT` with the help of the `VPFIT` front-end/back-end programs RDGEN and AUTOVP`IN`.

`VPFIT` decomposes segments of spectra into a set of Voigt profiles characterized by three parameters each: line redshift z_{abs}, Doppler parameter b, and column density $N_{\mathrm{H\,I}}$ for the hydrogen Lyα transition. We set up `VPFIT` to explore the range of parameters $1 \leq b/\mathrm{km\,s^{-1}} \leq 300$ and $11.5 \leq \log(N_{\mathrm{H\,I}}/\mathrm{cm^{-2}}) \leq 16$ when fitting absorption profiles. We chose to fit in this $N_{\mathrm{H\,I}}$ range in order to encompass typical optically thin Lyα absorbers ranging from low column densities (where most of the lines are comparable to noise) to very rare high column densities. Concerning the Doppler parameter, the chosen fitting region ranges from narrow absorbers, which are unphysical and have broadening comparable to the UVES/HIRES resolution element, to broad absorbers that are substantially broader than the typical absorber around the cutoff for all $\log T_0$ and γ combinations in our grid. This choice of fitting range is appropriate, as the probability of encountering absorbers close to the edge of our fitting range drops to nearly zero at this redshift.

`VPFIT` finds the best fit by varying the profile parameters and searching for a solution that minimizes the χ^2. If the χ^2 is not satisfying, then further absorption components are added until the fit converges or no longer improves. We take into account that `VPFIT` often has difficulty fitting the boundaries of spectra by artificially increasing the length of the sight lines. For this purpose we append the first (last) quarter of the spectra to the end (beginning) of it, therefore making the spectra longer by 50%. This manipulation does not cause discontinuities in the flux, as the simulation box is periodic. We later ignore absorbers within the artificially enlarged areas.

---

[2] VPFIT: http://www.ast.cam.ac.uk/~rfc/vpfit.html

Additionally, in order to avoid using badly constrained absorber parameters, we exclude points that have relative uncertainties worse than 50% in b or $N_{\mathrm{H\,I}}$. These lines are rejected in order to remove absorbers that are badly constrained. As discussed in § 2.1.3, most of these lines arise in blended and noisy regions. Additionally, as the log errors are proportional to the relative errors, we expect a 50% relative error to be $\mathrm{d}\log x = \ln(10)\cdot \mathrm{d}x/x = \ln(10)\cdot 0.5 \simeq 1.15$ (x being either $N_{\mathrm{H\,I}}$ or b). These uncertainties are substantially larger than our Kernel Density Estimation (KDE) bandwidth used in this study (see § 3.2.1) which additionally motivated us to to exclude these absorbers. Finally, filtering these lines consistently in data and models should not bias our results, as these are mostly `VPFIT` artifacts and will consistently arise whenever there is noise and blending.

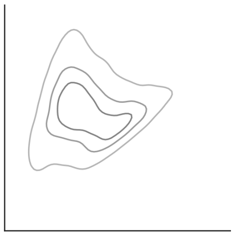
For every combination of $\log T_0$ and γ, a b-$N_{\mathrm{H\,I}}$ distribution can be generated from all absorbers found for all skewers. One example with $(\log T_0, \gamma) = (4.011, 1.333)$ is shown in the left panel of Figure 3.3.

3.2. Method for Emulating the Full b-$N_{\mathrm{H\,I}}$ Distribution

In this section, we introduce the method used to generate PDFs of b-$N_{\mathrm{H\,I}}$ distributions at any location in thermal parameter space based on our grid of simulated thermal models. For each thermal model, we perform KDE to determine $P(\log N_{\mathrm{H\,I}}, \log b)$ from the discrete absorbers identified by `VPFIT`. To interpolate the b-$N_{\mathrm{H\,I}}$ distribution between points in our parameter grid, we modified the emulation technique of Heitmann et al. (2006) and Habib et al. (2007), initially developed for power spectrum analysis, to our purpose. Note that this approach has also been used in the context of measurements of the evolution of the thermal state of the IGM in Rorai et al. (2013, 2017b) and Walther et al. (2019).

We apply principal component analysis (PCA) to decompose this set of probability distribution maps onto a set of basis vectors, yielding a set of coefficients $\Theta_j(T_0, \gamma)$ for each thermal model corresponding to principal component vectors $\boldsymbol{e}_j$. We then use Gaussian process (GP) interpolation to evaluate these coefficients at arbitrary locations in parameter space, which, combined with the basis vectors, results in a model for $P(\log N_{\mathrm{H\,I}}\,, \log b \mid \log T_0, \gamma)$.

Finally, we present a Bayesian method for determining the posterior distribution of thermal parameters from an observed set of $\log N_{\mathrm{H\,I}}$ and $\log b$. We refer to this procedure of model construction and inference, based on PCA decomposition of KDE estimates of a PDF, as the PKP method. The details of each step are discussed in what follows.



3.2.1. Kernel Density Estimation of the b-$N_{\mathrm{H\,I}}$ Distribution's Probability Distribution

In the first step of the PKP approach we use KDE to construct the probability density distribution from which points in the b-$N_{\mathrm{H\,I}}$ distributions of our models were drawn. This is achieved by treating each data point $\{\log N_{\mathrm{HI,i}}, \log b_i\}$ as a smooth kernel centered at the measurement positions $\log N_{\mathrm{HI,i}}$ and $\log b_i$. We use a Gaussian kernel of the form

$$K_i(\sigma_{\log N_{\mathrm{HI}}}, \sigma_{\log b}) = \frac{1}{2\pi\sigma_{\log N_{\mathrm{HI}}}\sigma_{\log b}} \times \exp\left(-\frac{1}{2}\left[\frac{(\log N_{\mathrm{HI}} - \log N_{\mathrm{HI,i}})^2}{\sigma^2_{\log N_{\mathrm{HI}}}} + \frac{(\log b - \log b_i)^2}{\sigma^2_{\log b}}\right]\right), \tag{3.3}$$

characterized by a bandwidth $(\sigma_{\log N_{\mathrm{HI}}}, \sigma_{\log b})$ that regulates how much one wishes to smooth a measurement in each dimension. Note that the kernel used in eqn. 3.3 assumes no correlation between $\log N_{\mathrm{HI,i}}$ and $\log b_i$ for a given pair. This assumption should not significantly affect the estimated PDFs, because the single kernels overlap substantially.

With every measurement described as a smooth distribution, we can generate an estimate for the probability density function from which a set of measurements $\{\log N_{\mathrm{HI,j}}\ , \log b_j\}$ with $j = 1, ..., N$, was drawn:

$$P(\log N_{\mathrm{HI}}, \log b) = \frac{1}{N}\sum_{j=1}^{N} K_j(\sigma_{\log N_{\mathrm{HI}}}, \sigma_{\log b}). \tag{3.4}$$

In other words, we compute $P(\log N_{\mathrm{H\,I}}, \log b)$ by replacing each measurement with a Gaussian kernel with a constant bandwidth, summing them up and normalizing the distribution.

In this study, we compute KDEs using the package `KDEMultivariate` from the `statsmodels` python module (Seabold & Perktold 2010). An example of this method applied to one of our b-$N_{\mathrm{H\,I}}$ distributions is shown in the right panel of Figure 3.3 for one particular combination of thermal parameters $(\log T_0, \gamma) = (4.011, 1.333)$, which can be compared to the points in the b-$N_{\mathrm{H\,I}}$ distribution determined by `VPFIT` in the left panel.

We generate KDE-based $P(\log N_{\mathrm{H\,I}}, \log b)$ for every thermal parameter combination in our standard thermal grid by applying KDE to the points in the b-$N_{\mathrm{H\,I}}$ distribution determined by `VPFIT`, using a bandwidth of $(\sigma_{\log N_{\mathrm{HI}}}, \sigma_{\log b}) = (0.08, 0.032)$ for each dimension. We tuned our bandwidth using mock datasets in order to avoid oversmoothing of $P(\log N_{\mathrm{H\,I}}\ , \log b \mid \log T_0, \gamma)$, which can wash out structure in the distribution. Additionally, oversmoothing shifts the peak of $P(\log N_{\mathrm{H\,I}}\ , \log b \mid \log T_0, \gamma)$ toward high

b owing to the asymmetry of the distribution, resulting in a distribution that has its maximum clearly shifted from the highest concentration of absorbers in the cloud of points used to generate it. At the same time we were careful not to undersmooth the distribution, which leads to a noisy PDF.

For comparison, a Silverman estimation of the optimal bandwidth (Silverman 1986) for our dataset, which assumes that the underlying distribution is Gaussian, typically yields a bandwidth of $(0.1, 0.04)$. This choice resulted in a very slight bias in our measurements for mock data in the context of the inference test described in § 3.3.2, indicating that this choice of bandwidth oversmoothes our distributions.

To illustrate the sensitivity of our PDF to thermal parameters, we show $P(\log N_{\mathrm{HI}}, \log b \mid \log T_0, \gamma)$ for different $\log T_0$ and γ combinations in Figure 1.8. We observe that, as expected, most of the sensitivity of the b-N_{HI} distribution with respect to the parameters of the TDR lies in its lower b envelope. Therefore, in the limit of a measurement of T_0 and γ, our approach can be interpreted as an alternative way of retrieving the cutoff (although without many of the problems associated with iterative cutoff fitting as discussed in § 2.4.6). Nevertheless, our method can be expanded to any changes in the general form of the b-N_{HI} distribution, provided that these are properly modeled in the simulations. The example of T_0 and γ is an interesting starting point to apply our method to, but should not be seen as its sole application. We know, for instance, that λ_P (Garzilli et al. 2015, 2018), the fraction of the gas in the warm-hot phase (Danforth et al. 2016) and galactic feedback (Viel et al. 2017) affect the shape of the b-N_{HI} distribution above the location of the cutoff. In principle, our method should be sensitive to these parameters as well.

For better intuition about the thermal sensitivity of the b-N_{HI} distribution we also added Figures, constructed from the output of hydrodynamic simulations described in § 3.4.1, to appendix B.1. These can be be viewed as animations in the HTML version of this manuscript.

3.2.2. Decomposition of the PDF into Principal Components

Given the nonparametric nature of KDE, there is no direct way to generate $P(\log N_{\mathrm{HI}}, \log b \mid \log T_0, \gamma)$ for combinations of $\log T_0$ and γ between points in our thermal grid positions. For this to be possible, we have to parameterize the $P(\log N_{\mathrm{HI}},\log b \mid \log T_0, \gamma)$ maps. To this end, we evaluate the KDE of each b-N_{HI} distribution on a 100×100 mesh in the b-N_{HI} plane and then decompose these pixelized PDFs onto a set of linear independent principal components, thus parameterizing the KDE-based $P(\log N_{\mathrm{HI}}, \log b \mid \log T_0, \gamma)$ with PCA coefficients and a set of basis vectors.

Specifically, we discretized the PDFs in the region $11.5 \leq \log(N_{\mathrm{HI}}/\mathrm{cm}^{-2}) \leq 16.$ and $0.8 \leq \log(b/\mathrm{km\,s}^{-1}) \leq 2.2$), adopting a pixel size $(0.04, 0.014)$ in $(\log N_{\mathrm{HI}}, \log b)$, which is a factor of 2 smaller than the bandwidth chosen for the KDE. Then, we compute the (natural) logarithm of the probabilities at every pixel. Given our small pixels, we expect no significant change in the shape of the b-$N_{\mathrm{H\,I}}$ distribution due to pixelization. All examples of smooth b-$N_{\mathrm{H\,I}}$ distributions shown in this chapter and in chapter 1 are pixelized on this grid (see e.g. Figures 3.3, 1.8, and 3.5).

The PCA is performed by decomposing our discrete maps into a basis of principal component vectors $\boldsymbol{e}_j$, which makes it possible to recover any model in our grid by linearly combining the principal component vectors, using the coefficients $\Theta_j(\log T_0, \gamma)$ and adding them to the mean map $\boldsymbol{\mu}(N_{\mathrm{H\,I}}, b)$:

$$\begin{aligned} \ln P(\log N_{\mathrm{H\,I}}, \log b | \log T_0, \gamma) &= \boldsymbol{\mu}(\log N_{\mathrm{H\,I}}, \log b) \\ &+ \sum_{j=1}^{N} \Theta_j(\log T_0, \gamma) \boldsymbol{e}_j(\log N_{\mathrm{H\,I}}, \log b), \end{aligned} \tag{3.5}$$

where N is the number of models available, in this case $N = 100$, and the components are ranked by their contribution to the cumulative variance of the dataset. In short, the PCA decomposes a matrix of all vectorized $\ln P(\log N_{\mathrm{H\,I}}, \log b)$ maps into a basis of 100 principal component vectors with 100 coefficients each.

In Figure 3.4 we show the $\boldsymbol{\mu}(\log N_{\mathrm{H\,I}}, \log b)$ map and the first three principal component vectors (reshaped to an image of 100×100 pixels) and coefficients from our analysis. Note that PCA is a standard method for dimensionality reduction, as it allows one to choose the principal components that encompass most of the variance within the data by ignoring components that do not contribute substantially to the cumulative variance. The cumulative contribution to the total variance is computed by first dividing the eigenvalues from the singular value decomposition method used in the PCA by their sum, ordering them in descending order, and computing their cumulative sum. For illustration, the first three components shown in Figure 3.4 already account for 83.1% of the cumulative variance in the models. At present, we are not interested in dimensionality reduction, and keeping all 100 PCA components is not computationally prohibitive for the current case. By PCA-decomposing the KDEs in our grid, we are simply describing each of the discretized $P(\log N_{\mathrm{H\,I}}\,, \log b \mid \log T_0, \gamma)$ with a set of coefficients $\Theta_j(T_0, \gamma)$ and basis vectors, enabling a parametric description of $P(\log N_{\mathrm{H\,I}}\,, \log b \mid \log T_0, \gamma)$.

There are two reasons why we carried out the PCA on $\ln P(\log N_{\mathrm{H\,I}}, \log b)$: first, because we will interpolate PCA components of $\ln P(\log N_{\mathrm{H\,I}}\,, \log b | \log T_0, \gamma)$ maps (§ 3.2.3) in thermal parameter space, and these PDFs have sharp features (such as the low b-cutoff). Computing the natural logarithm is desirable to reduce interpolation errors; second, we do this for a practical reason, as we will ultimately tie this analysis to a Markov chain Monte Carlo (MCMC) algorithm that works with the log-likelihood.

The disadvantage of working with the natural logarithm of $P(\log N_{\mathrm{H\,I}}, \log b \mid \log T_0, \gamma)$ is that the probability fluctuations around zero are amplified, which can destabilize the

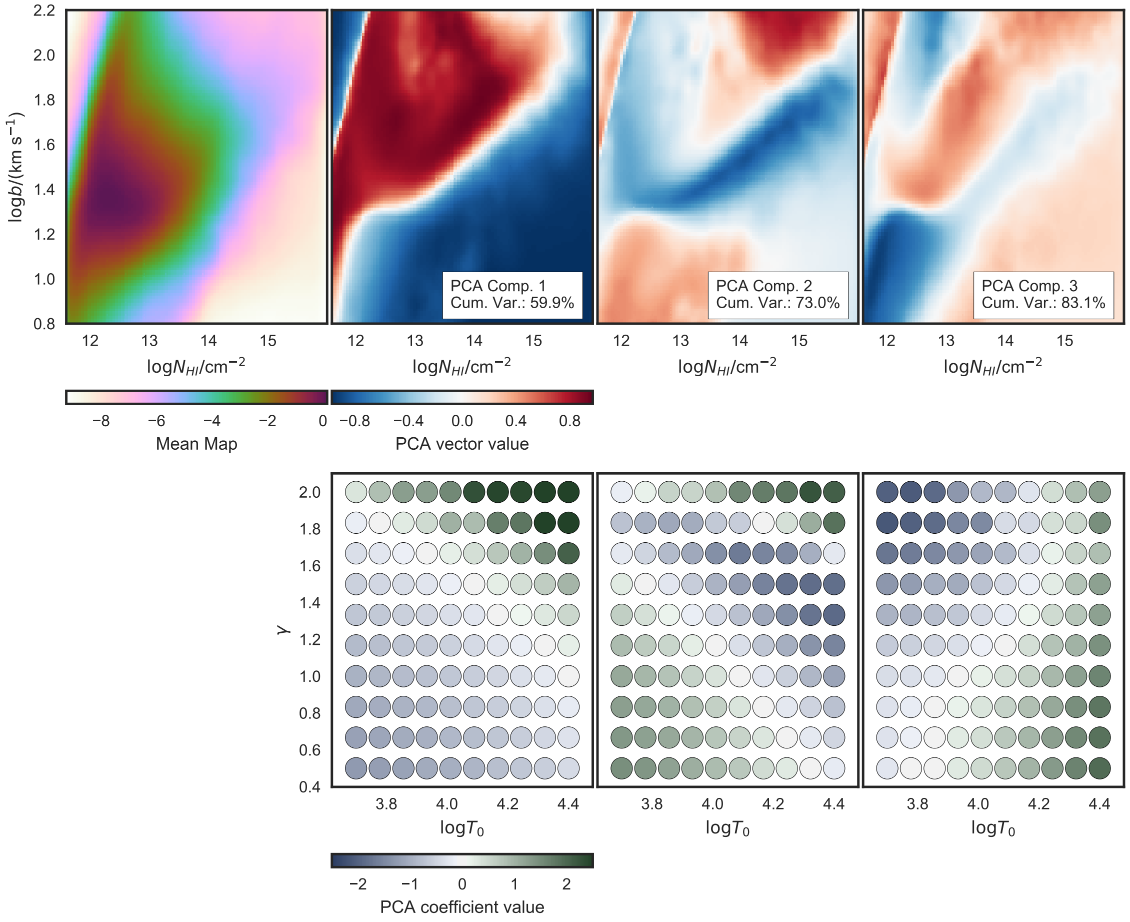


Figure 3.4. Top row: The mean map $\boldsymbol{\mu}(\log N_{\mathrm{HI}}, \log b)$ and the first three principal component vectors $\boldsymbol{e}_j$ from our PCA of our model maps. Note that the decomposition was carried out in the natural logarithm of the probability. The vectors were reshaped to the map form of 100 × 100 pixels and are sorted by contribution to the cumulative variance (see text for details). **Bottom row:** The corresponding principal component coefficients $\Theta_j(\log T_0, \gamma)$ for each map.

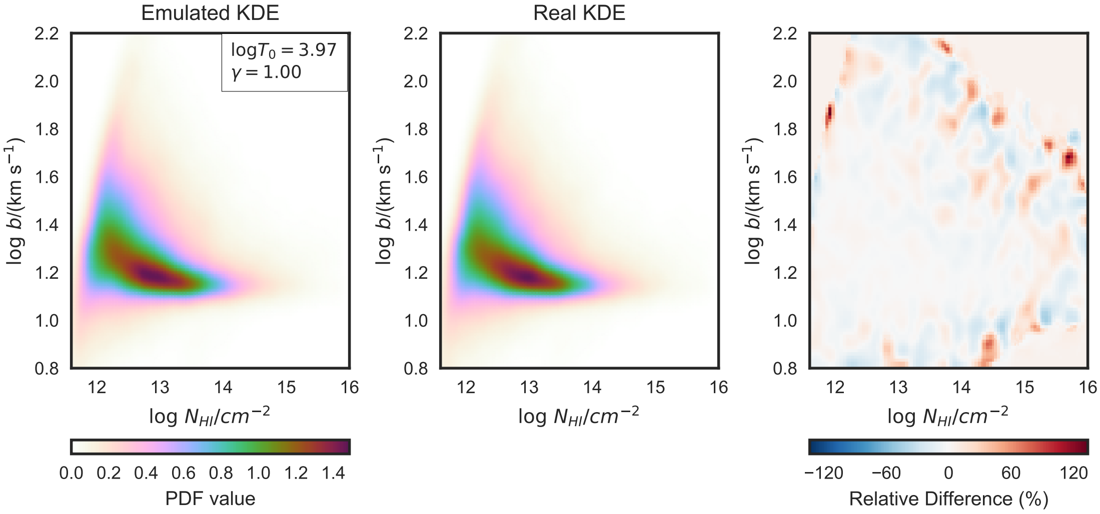


Figure 3.5. Comparison of interpolated and measured $P(\log N_{\mathrm{HI}}\,, \log b \mid \log T_0, \gamma)$ for a model in our test grid, i.e. not included in the grid used for constructing the $\ln P(\log N_{\mathrm{HI}}\,, \log b|\log T_0, \gamma)$ emulator. Thermal parameters are $(\log T_0, \gamma) = (3.972, 1.0)$. **Left:** $P(\log N_{\mathrm{HI}}\,, \log b \mid \log T_0, \gamma)$ constructed by interpolating PCA components using GP interpolation. **Middle:** $P(\log N_{\mathrm{HI}}\,, \log b \mid \log T_0, \gamma)$ generated from KDE of the PDF directly from the `VPFIT` output at the same thermal parameters. **Right:** the difference of emulated and original b-N_{HI} distribution relative to the original b-N_{HI} distribution illustrates that we are able to accurately emulate the PDF between our grid points. The fact that we see no relative difference in the edges of the rightmost diagram comes from the fact that we set a density threshold under which the probability was set to zero (see § 3.2.2).

interpolation process in the low-probability regions. To avoid interpolation artifacts in the low-probability regime, we simply apply a probability threshold to all our discrete $\ln P(\log N_{\mathrm{HI}}\,, \log b|\log T_0, \gamma)$ maps under which all probabilities are set to zero. We chose to set this threshold at the value of the 20th percentile of the probability values for each map. Typically this threshold corresponds to a probability < 0.003, i.e. it only affects the lowest probabilities of $P(\log N_{\mathrm{HI}}, \log b \mid \log T_0, \gamma)$ and does not vary strongly from model to model. Varying this threshold did not affect our emulated distributions substantially for values lower than the 40th percentile of the probability values for each map, as the cut involves the lowest-probability regions.

3.2.3. Emulating the PDF

Finally, we train a GP on the PCA coefficients for our discrete model grid (using GEORGE Ambikasaran et al. 2016). This allows us to generate $\ln P(\log N_{\mathrm{HI}}, \log b)$ at arbitrary $\log T_0$ and γ combinations.

A GP is basically a stochastic process for which every finite subset of random variables is normally distributed, i.e. it can be fully described by its mean and a covariance function. The covariance function is a measure of how much two points in parameter space $\boldsymbol{\vartheta}_l$ and $\boldsymbol{\vartheta}_m$ are covariant, $\boldsymbol{\vartheta}$ being a vector with $(\log T_0, \gamma)$ in our parameter space. We adopt a standard choice for the covariance C, which is a squared-exponential kernel plus an additional white-noise contribution, with the form

$$C(\boldsymbol{\vartheta}_l, \boldsymbol{\vartheta}_m) = \exp\left(-0.5\,(\boldsymbol{\vartheta_l} - \boldsymbol{\vartheta_m})\,C_h^{-1}\,(\boldsymbol{\vartheta_l} - \boldsymbol{\vartheta_m})\right) + \sigma_n \delta_{\mathrm{lm}}, \tag{3.6}$$

where C_h is chosen to be a diagonal matrix with a smoothing length h_l for every dimension, i.e. the characteristic distance beyond which the covariance between two points drops, and σ_n parameterizes the white-noise term. We chose h_l to be a constant with the value of 20% of our standard thermal grid length in each dimension[3] (larger than the typical grid separation). This guarantees that the interpolation will correlate coefficients $\Theta_j(T_{0,i}, \gamma_i)$ from neighboring points in the grid.

There are an infinite number of functions that satisfy a GP with a specific mean and covariance, but the interpolation (or regression) part comes in once we only select the subset of functions that are constrained to pass through a particular set of points. In our case, we have a vector of 100 PCA coefficients $\Theta_j(T_{0,i}, \gamma_i)$ for each model combination i in our grid of 100 simulations. Although GP interpolation can be generalized for the case in which the computed PCA coefficients have uncertainties by having the white-noise term $\sigma_n \delta_{\mathrm{ij}}$ in eqn. 3.6, we decided to assume that these PCA coefficients have no uncertainty, i.e. we force the interpolation to pass nearly perfectly through the measured $\Theta_j(T_{0,i}, \gamma_i)$ by setting σ_n to nearly zero[4]. This means that our emulator essentially recovers the b-N_{HI} distribution maps perfectly at the thermal grid positions.

We illustrate the accuracy of our procedure in Figure 3.5. In the left panel we show an emulated $P(\log N_{\mathrm{HI}}\,, \log b \mid \log T_0, \gamma)$ for a $(\log T_0, \gamma) = (3.972, 1.0)$ combination between points in our standard grid. The middle panel shows the true KDE-based PDF from the `VPFIT` output for this thermal model (taken from our test grid). The right panel shows the relative difference between the two PDFs, which scatters around 0 and is typically of the order of 3% in probability in the high-probability regions, indicating that we can successfully interpolate between models. The difference drops to zero in the far edges owing to the thresholding of the density described in § 3.2.2. There are some peaks in the relative difference close to the edges that arise simply because the 20th percentile density thresholding did not affect the exact same pixels in the real vs. the emulated distribution.

---

[3] More specifically, prior to the interpolation, our thermal grid was renormalized to the range 0-1 in each dimension and a kernel size of 0.2 was used. See appendix B.2.1 for a motivation of this choice.

[4] The emulation would not converge when setting $\sigma_n = 0$, so we adopted the default `TINY` noise value 1.25×10^{-12} from the GEORGE library.

We will further discuss the effect of the GP interpolation when performing mock measurements in § 3.3.2.

3.2.4. Parameter Inference

We use the $\ln P(\log N_{\mathrm{H\,I}}, \log b | \log T_0, \gamma)$ emulator as a basis for calculating the likelihood of a dataset given model parameters. The probability of measuring a single absorption line $(N_{\mathrm{H\,I},i}, b_i)$ is given by the PDF $P(\log N_{\mathrm{H\,I}}, \log b \mid \log T_0, \gamma)$. Thus, the likelihood for measuring a set of N absorption lines $\log N_{\mathrm{HI}}, \log b$ is

$$\mathcal{L} = \prod_i^N P(\log N_{\mathrm{H\,I},i}, \log b_i | \log T_0, \gamma), \tag{3.7}$$

or in terms of log-likelihood

$$\ln \mathcal{L} = \sum_i^N \ln P(\log N_{\mathrm{H\,I},i}, \log b_i | \log T_0, \gamma). \tag{3.8}$$

Given that our emulator is able to generate model PDFs at any given point within the thermal parameter grid, we simply couple this log-likelihood to an MCMC algorithm to perform Bayesian inference of the model parameters. For this purpose we use the python package `emcee` (Foreman-Mackey et al. 2013) which implements the affine-invariant sampling technique (Goodman & Weare 2010). We assumed flat priors for both parameters that are truncated at the edges of our standard thermal grid for all MCMC runs presented in this study.

The key assumption of the likelihood above is that we treat the Lyα forest as being an uncorrelated distribution of lines such that we can look upon each $\log N_{\mathrm{H\,I}}$, $\log b$ measurement as a random draw from $P(\log N_{\mathrm{H\,I}}, \log b \mid \log T_0, \gamma)$. We expect that this assumption does not affect our likelihood substantially given the low level of spatial correlations in the Lyα forest (McDonald et al. 2006). We will carry out an inference test in § 3.3.2 and asses wether this affects mock measurements carried out with the PKP method.

3.3. Testing the Robustness of Our Inference

In this section we test the PKP method by carrying out mock measurements of $\log T_0$ and γ using MCMC. First, we show one example of a measurement, and then we test the robustness of our method by examining how the MCMC posteriors behave for measurements based on many random realizations of mock datasets for the models in our test grid.

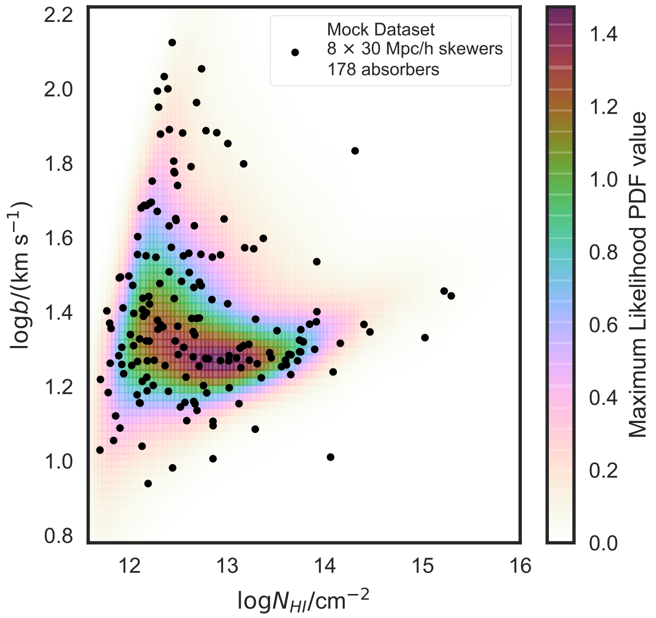


Figure 3.6. A mock realization of a data b-N_{HI} distribution composed of the absorbers from eight randomly chosen skewers from a simulation with $(\log T_0, \gamma) = (4.050, 1.333)$. An emulated $P(\log N_{\mathrm{HI}}, \log b \mid \log T_0, \gamma)$ based on the median values of the marginal distributions of the corresponding MCMC posterior ($\log T_0 = 4.054$ and $\gamma = 1.303$, see Figure 3.7) is shown for comparison. Figure credit: Hiss et al. (2019).

3.3.1. Measurement Example

As an example of a mock measurement we select the absorbers from a sample of eight random skewers extracted from a model with $(\log T_0, \gamma) = (4.050, 1.333)$ in our test grid (the blue points in Figure 3.2). The corresponding dataset is shown as black points in Figure 3.6. For reference, this mock dataset is comparable in terms of path length to the redshift range 1.9-2.1 provided by a single QSO spectrum in the analysis presented in § 2. Specifically, this dataset is generated from a pathlength of 240Mpc. While a single Lyα forest at this redshift (between Lyα and Lyβ emission peaks) covers $\sim$ 620Mpc (from $z = 2.1$ to 1.7), the redshift bin used in § 2.1 ranged from 1.9-2.1, so each QSO contributed $\sim$ 295Mpc. Effectively, due to the masking applied to the data in order to filter possible metal contaminants and the path-length reduction associated with it, our mock dataset corresponds to nearly two sight lines in terms of number of absorbers at this redshift range.

The results of our MCMC inference for this particular mock dataset are shown in Figure 3.7. We observe the well-known strong degeneracy in the measurement of $\log T_0$ and γ, which is a result from setting the pivot point of the TDR at mean density (see

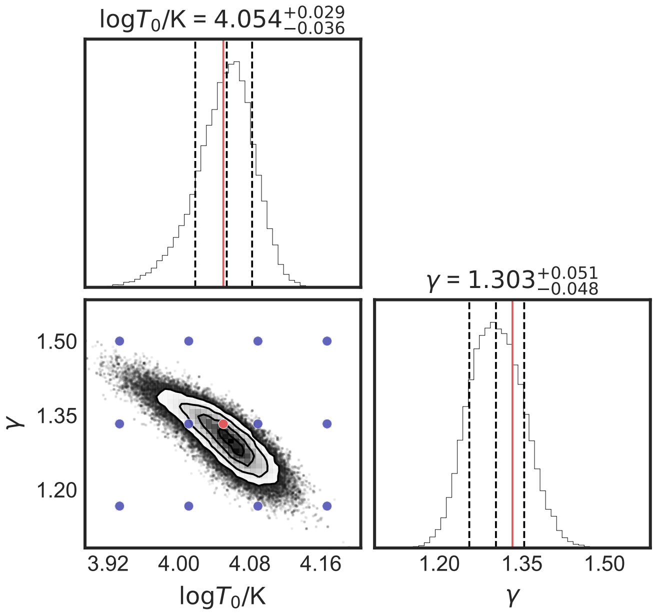


Figure 3.7. MCMC posterior for a mock dataset composed of eight randomly chosen skewers (absorbers shown in Figure 3.6) extracted from a model in our test grid with thermal parameters shown in red. A zoomed-in part of the thermal grid used for constructing the emulator on which this measurement is based is shown in blue. The model from which the mock data were chosen (red circle) is not included when constructing the $\ln P(\log N_{\mathrm{HI}}, \log b|\log T_0, \gamma)$ emulator.Figure credit: Hiss et al. (2019).

e.g. Lidz et al. 2010; Becker et al. 2011; Walther et al. 2019; Hiss et al. 2018). We obtain $\log T_0 = 4.054^{+0.029}_{-0.036}$ and $\gamma = 1.303^{+0.051}_{-0.048}$, whereby the errors are calculated based on the 16th and 84th percentiles of the marginalized distributions of the MCMC posterior. One observes that this is remarkably close to the true model that the dataset was drawn from (indicated by the red circle and lines in Figure 3.7). We can illustrate the inferred model PDF by inputting these measured thermal parameters (i.e. the median of the individual marginalized posteriors) into into our emulator, retrieving the corresponding $\ln P(\log N_{\mathrm{HI}}, \log b|\log T_0, \gamma)$ and computing $\exp(\ln P(\log N_{\mathrm{HI}}, \log b|\log T_0, \gamma))$, which is shown by the color-coded distribution in Figure 3.6.

3.3.2. Inference Test

In order to further test the robustness of our method, we perform measurements of $\log T_0$ and γ using 10 mock data realizations of b-N_{HI} distributions (based on eight random skewers each) for each of the 16 models in the test grid. Our uncertainties are quantified based on the two-dimensional MCMC posteriors (see e.g. Figure 3.7).

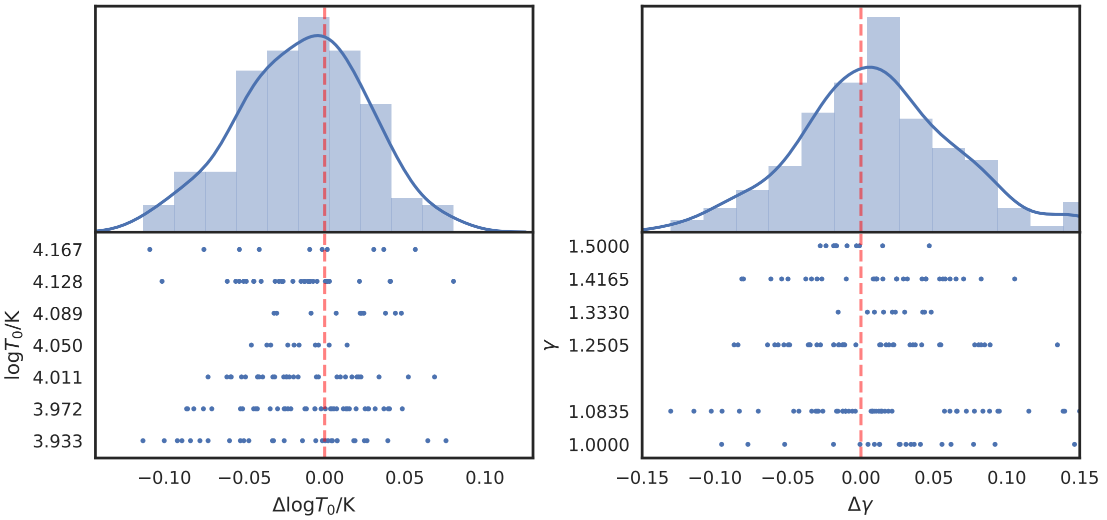


Figure 3.8. Distribution of the difference between the true values and the medians of MCMC posteriors for 10 random realizations of mock datasets with eight skewers each for all the 16 test models (blue in Figure 3.2). The likelihoods for these measurements were calculated based on $P(\log N_{\rm HI}\,, \log b \mid \log T_0, \gamma)$ generated using our emulator, which did not include these models. The differences between true values and MCMC-based estimates are shown as blue points in the bottom panels for each realization. A histogram of all measurements put together is shown in blue, while the blue line corresponds to a 1D KDE of the differences in the histogram. The red dashed line illustrates a perfect measurement. Figure credit: Hiss et al. (2019).

Testing our measurements by inspecting many realizations of mock datasets will reveal wether our method is returning valid posterior probability distributions.

Given that we are dealing with models exactly between our standard grid points, this test will show wether interpolation errors in $\ln P(\log N_{\rm HI}\,, \log b|\log T_0, \gamma)$ result in biased measurements. This is a crucial test given that our typical MCMC contours have uncertainties that are comparable to the characteristic separation between models in our thermal grid, which is illustrated by the blue grid points shown in Figure 3.7. Furthermore, an inference test will fail, for instance, if our assumption that we can neglect spatial correlations in the Lyα forest in the likelihood in eqn. 3.8 is incorrect.

We test wether the uncertainties derived from the MCMC posteriors are sensible by carrying out the following exercise. For all of the 160 posteriors, i.e. 16 distinct models times 10 mock realizations of each model, we quantify how often the true values of the thermal parameters used land within the 68% and 95% confidence regions of the corresponding 2D MCMC posterior. We observe that the true values are within the 68% confidence region 68.7% (110/160) of the time and that they are within the 95%

confidence region 96.9% (155/160) of the time. This convincingly indicates that our posterior distributions are robust and that we are not over- or underestimating our uncertainties.

As a further test of whether our inference is significantly biased, we examine the distribution of the difference between the true values of $\log T_0$ and γ and the median of the marginalized distributions of the MCMC posteriors: $\Delta \log T_0 = \log T_{0,\mathrm{true}} - \log T_{0,\mathrm{MCMC}}$ and $\Delta\gamma = \gamma_{\mathrm{true}} - \gamma_{\mathrm{MCMC}}$. The distributions of these differences are presented in Figure 3.8. We see that the distributions are centered around zero, indicating that any bias associated with our method is smaller than the resulting uncertainties. Note that in this initial experiment we are deliberately only carrying out our tests for the measurement of T_0 and γ, not taking into account the correlations with other parameters such as pressure smoothing scale λ_p or amplitude of the UVB. While certainly important, adding these dimensions to our analysis is beyond the scope of introducing and testing our new approach.

3.4. Pilot Study: A Measurement of Thermal Parameters at z=2

The DM-only models used for our inference test in § 3.3.2 use an approximation for generating flux skewers that does not capture the full physical picture necessary to properly represent the IGM (Sorini et al. 2016). While DM-only simulations were sufficient for our initial tests (see § 3.1), for a realistic measurement involving real observational data, one has to use hydrodynamic simulations to generate model distributions. In this section we apply the PKP method to real Lyα forest absorption-line data using a grid of hydrodynamic simulations to model $P(\log N_{\mathrm{H\,I}}, \log b \mid \log T_0, \gamma)$.

3.4.1. The b-$N_{\mathrm{H\,I}}$ Distribution Conditional Probability Density Function from Hydrodynamic Simulations

Following the approach described in § 3.1 and 3.2, we now generate models of $P(\log N_{\mathrm{H\,I}}, \log b \mid \log T_0, \gamma)$ by applying `VPFIT` to simulated skewers drawn from hydrodynamic simulations of different thermal models. Hydrodynamic simulations provide the general physical conditions that give rise to the Lyα forest directly from first principles, with the exception of reionization effects, thus resulting in realistic b-$N_{\mathrm{H\,I}}$ distributions. Additionally, pressure smoothing of absorbers is accounted for in a physical way as opposed to the artificial smoothing of the density field that was used in the DM models. The disadvantage associated with hydrodynamic simulations is that, unlike the DM-based model, it is costly to generate large grids in T_0 and γ at a given redshift, which could pose a problem given the high precision our method can achieve. Nevertheless, grids of ~ 30 hydrodynamic simulations are computationally feasible (see below).

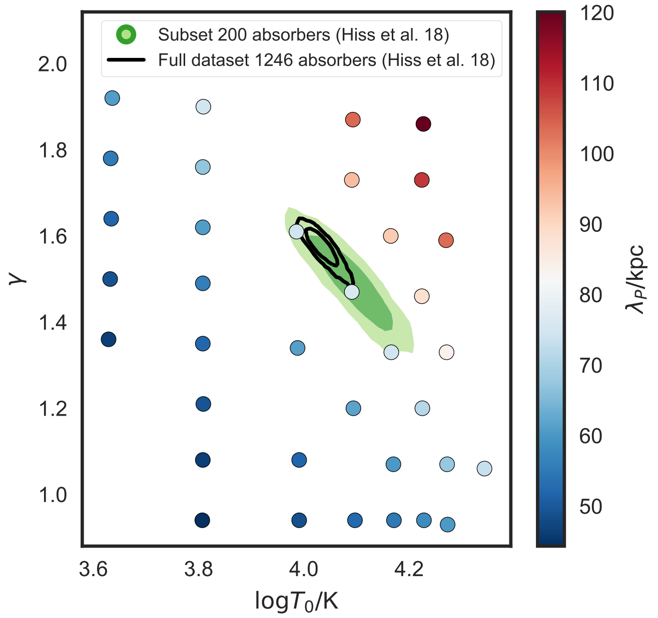


Figure 3.9. Thermal grid from snapshots of hydrodynamic simulations from the THERMAL suite at $z = 2$ used in this pilot study. The points are colored based on the pressure smoothing scale λ_P. For comparison with the characteristic grid separation, we show the measurements (see § 3.4.3) achieved using the full dataset at $z = 2$ shown in § 2.1.3 (black contour lines) and a subset of 200 absorbers from this dataset (green contours). Figure credit: Hiss et al. (2019).

For the purpose of generating a basis of model b-N_{HI} distributions, we use part of the publicly available THERMAL[5] suite of Nyx simulations (Almgren et al. 2013; Lukić et al. 2015) presented in § 2.2. The THERMAL suite consists of more than 60 Nyx hydrodynamic simulations with different thermal histories and $L_{\mathrm{box}} = 20\,\mathrm{Mpc}/h$ and 1024^3 cells based on a Planck Collaboration et al. (2014) cosmology $\Omega_m = 0.3192$, $\Omega_\Lambda = 0.6808$, $\Omega_b = 0.04964$, $h = 0.6704$, $n_s = 0.96$, $\sigma_8 = 0.826$. We chose a grid consisting of a subset of 36 simulation snapshots at $z = 2$ with different combinations of T_0, γ and λ_P that result from different thermal evolutions (Oñorbe et al. 2017), shown in Figure 3.9.

Note that although arbitrary λ_P values could be generated in principle, it would require substantial computing power to fine-tune the reionization histories to do so. As discussed in Walther et al. (2019), it is difficult to generate physically realistic

---

[5] Url: `http://thermal.joseonorbe.com/`

models without correlating the TDR parameters and λ_P, because the pressure smoothing scale depends on the integrated thermal history of the IGM. Due to computing time restrictions, we generate only physically motivated λ_P that are correlated with the TDR parameters, i.e. high (low) T_0 and γ combinations generate large (small) values of λ_P.

Following our discussion in § 3.2.1, we apply the same KDE procedure to the `VPFIT` output of our simulations and then construct a $\ln P(\log N_{\rm HI}, \log b|\log T_0, \gamma)$ emulator based on simulated b-$N_{\rm HI}$ distributions (as in § 3.2.3). For the $\ln P(\log N_{\rm HI}, \log b|\log T_0, \gamma)$ emulation we apply the same PCA and GP interpolation scheme, adopting smoothing lengths h in the covariance (see eqn. 3.6) for the interpolator that is 50% of the grid size in the $\log T_0$ direction and 20% in the γ direction. Additionally, for the white-noise term in eqn. 3.6 we chose $\sigma_n = 0.01$, which allows for small deviations in the interpolation at the grid points. These changes relative to the DM only emulation were arrived at via visual inspection of the emulated PDFs. Specifically, we changed these parameters until no interpolation artifacts were present throughout the grid. A motivation of this choice of white noise contribution is presented in the appendix B.2.2. Additionally, similar to the analysis of mock datasets in § 3.3.2, we checked wether we accurately recover the thermal parameters at the grid positions and found that the results were unbiased. This indicates that the different GP smoothing parameters and white-noise term added when using hydrodynamic simulations do not significantly bias our inference.

3.4.2. Absorption-line Dataset

In order to carry out a measurement, we use the absorption-line data presented in § 2.1.3 which consists of 1246 absorption lines[6] at $1.9 \leq z < 2.1$.

One problem that could bias the results of our method are outliers with low b in the b-$N_{\rm HI}$ distribution. In § 2.1.5 we argue that these are narrow lines added by `VPFIT` in order to decrease the χ^2 of the fit in blended absorption features and unidentified metal absorbers wrongly assumed to be Lyα lines (as observed by Schaye et al. (1999); Rudie et al. (2012a)). Blending artifacts should not have a severe impact on our measurements, as a proper forward modeling of the simulated sight lines should include the same sort of contamination in our model $P(\log N_{\rm HI}, \log b \mid \log T_0, \gamma)$.

As for dealing with metal line contamination, the dataset used was carefully masked for metal absorption systems, as described in § 2.1.4. The severity of metal line contamination is strongly redshift dependent, as the identification of metal absorbers in the Lyα forest becomes increasingly difficult at higher redshift (and nearly unfeasible at $z \gtrsim 3.5$) owing to line blanketing as the effective optical depth of the Lyα forest

---

[6] In line with the approach presented presented in § 2.1.3, we excluded absorbers that have relative uncertainties worse than 50% in b or $N_{\rm HI}$ from the observational dataset. For consistency, the same recipe was applied to the lines of sight extracted from our hydrodynamic simulations.

increases. In our case, the contamination should be relatively mild, given that metal line absorbers are more easily identified at lower redshifts and that these data were previously masked for potential contaminants using different automatic and interactive techniques. Nevertheless, there are remaining unidentified contaminants that have to be excluded with some sort of outlier rejection.

As in § 2.1.5 we implement an iterative 2σ rejection procedure based on Rudie et al. (2012a) that rejects potential narrow-line contaminants in the range $12.5 < \log N_{\mathrm{HI}}/\mathrm{cm}^{-2} < 14.5$. For simplicity, we decided to extrapolate the 2σ rejection line calculated in § 2.1.5 to the region $11.5 < \log N_{\mathrm{HI}}/\mathrm{cm}^{-2} < 16$ (shown as a gray dashed line in Figure 3.10) and discard all absorbers with $\log b$ lower than this line. Alternatively, one could implement a more elegant outlier modeling method such as the one used by Telikova et al. (2018), but here we opt for this simpler approach.

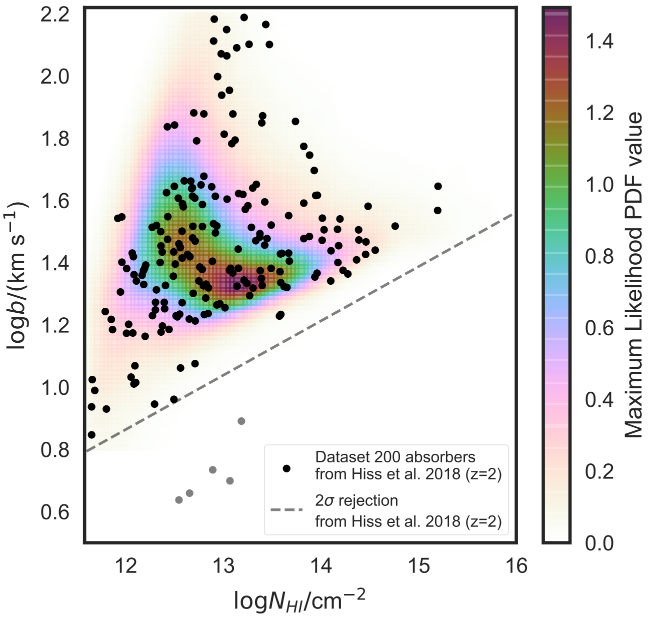


Figure 3.10. Subset of the b-N_{HI} distribution at $z = 2$ presented in § 2.1.3 composed of 200 randomly chosen absorbers (black points). To avoid possible narrow-line contaminants (gray points), only absorbers with b above the extrapolated 2σ rejection line from § 2.1.5 were chosen (gray dashed line). An emulated $P(\log N_{\mathrm{HI}}, \log b \mid \log T_0, \gamma)$ based on the median values of the marginal distributions of the corresponding MCMC posterior is shown for comparison. Figure credit: Hiss et al. (2019).

The $z = 2$ dataset from § 2.1.3 has a size of 1246 absorbers, and we have intuition from § 3.3.2 that this dataset size would result in percent-level precision, i.e. smaller than the spacing between our thermal grid points, making our inference susceptible to interpolation uncertainties. We thus decided to randomly choose a set of 200 absorbers from this dataset, hence with a similar number of lines to the mock dataset of eight skewers described in § 3.3. In contrast to the SNR modeling done in § 3.1.2 using a constant value of 63 per 6 km s^{-1}, we randomly chose the SNR from the real sight lines for the mock spectra from hydrodynamic simulations to better represent the noise distribution within the data (exactly as was done in § 2.2.3). Because of this approach, it makes more sense to chose a random subset of absorbers rather than selecting a random subset of QSO sight lines. For a discussion about how our results differ if we randomly choose QSO sight lines instead of absorbers, please refer to the appendix B.3.

To understand how our uncertainties compare to the typical separations between points in our thermal parameter grid, we show two sets of $\log T_0$-γ measurements in Figure 3.9. We will explain in detail how these contours were measured in the next section. But for the sake of the current discussion, note that the green contours result from analyzing a dataset of 200 absorbers, resulting in a precision comparable to our characteristic grid separation, whereas the black contours show a measurement using the complete dataset of 1246 absorbers. Clearly, using the full dataset results in an uncertainty substantially smaller than our grid spacing, which indicates that interpolation errors could be a significant issue. Given the exquisite precision delivered by the PKP method and the size of existing datasets, it is challenging to generate a grid of hydro simulations fine enough to do justice to the implied precision. Nevertheless, we believe that this is computationally within reach and will enable measurements of the thermal state of the IGM with unprecedented precision.

3.4.3. Results

In order to measure $\log T_0$ and γ, we carry out the same Bayesian measurement as described in § 3.3, this time using real data combined with $P(\log N_{\rm H\,I}, \log b)$ emulated from hydrodynamic simulations. The subset of 200 absorbers from § 2.1.3 are shown as black points in Figure 3.10, whereas the five gray points are the corresponding fraction of absorbers that are rejected. The green contours in Figure 3.9 show the MCMC posterior resulting from analyzing these data, from which we measure $\log T_0 = 4.092^{+0.050}_{-0.055}$ and $\gamma = 1.49^{+0.073}_{-0.074}$, whereby the errors are calculated based on the 16th and 84th percentiles of the marginalized distributions. We explore how this inference behaves for different random realizations of 200 absorbers in the appendix B.3. As before, we emulate the $P(\log N_{\rm H\,I}, \log b \mid \log T_0, \gamma)$ at these measured values, which is shown as the color-coded distribution in Figure 3.10.

Additionally, we carried out the same measurement using the full dataset of 1264 lines presented in § 2.1.3. As discussed in § 3.4.2, due to the current separations in our model grid, we have concerns about interpolation error at such a high level of precision.

Nevertheless, we wanted to illustrate the kind of precision achievable using existing data. With these caveats, we measure $\log T_0 = 4.034^{+0.022}_{-0.019}$ and $\gamma = 1.576^{+0.026}_{-0.032}$. The corresponding contours are shown in black in Figure 3.9. Importantly, compared to the measurement using a subset of these data, the uncertainties are smaller by a factor of approximately $\sqrt{6}$, which is the expected scaling due to the relative sizes of the datasets.

Comparison with Cutoff Fitting Results

These PKP-based results can be compared to the measurements presented in § 2.4 from the same dataset using the cutoff fitting approach. From the marginalized distributions of the Monte Carlo-based posteriors presented in § 2.4, we measured $\log T_0 = 4.137^{+0.050}_{-0.074}$ and $\gamma = 1.47^{+0.12}_{-0.10}$ at $z = 2$ using the 1264 Lyα absorbers.

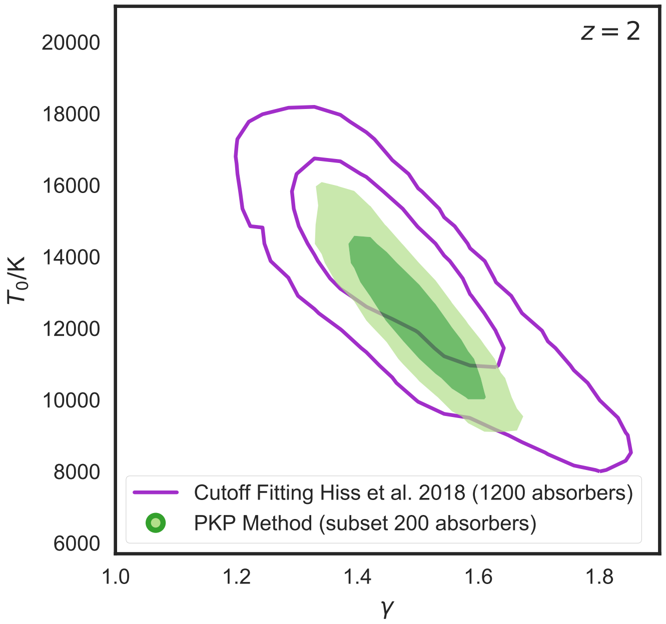


Figure 3.11. Comparison of the thermal parameter constraints from § 2 using the cutoff fitting method (purple contour lines) and our measurement using the PKP method (green contours). While the original dataset presented in § 2.4 has a size of 1246 absorbers, only 845 are actually used for cutoff fitting as a result of the fact that only absorbers with $12.5 \leq \log(N_{\mathrm{HI}}/\mathrm{cm}^{-2}) \leq 14.5$ and $8 < b/\mathrm{km\,s}^{-1} < 100$ are used. The cutoff fitting results are shown as purple contour lines. When using the PKP method described in this study, we achieve higher precision (green contours) while using a random subset of 200 absorbers from their data. Figure credit: Hiss et al. (2019).

As stated in § 3.4.3, when applying our new method to a subset of 200 absorbers from their dataset, we measure $\log T_0 = 4.092^{+0.050}_{-0.055}$ and $\gamma = 1.49^{+0.073}_{-0.074}$. In Figure 3.11 we compare our PKP-based measurement using just 200 absorbers from § 2.1.3 at $z = 2$ (green shaded contours) to the cutoff fitting measurement from § 2.4 (Figure 2.19) using the full dataset (purple contours).

A direct comparison of these measurements based on the size of the dataset used is challenging because both methods use different cuts in the data. While we use all absorbers within the allowed fitting range, the cutoff fitting method only uses the absorbers within $12.5 \leq \log(N_{\mathrm{H\,I}}/\mathrm{cm}^{-2}) \leq 14.5$ and $8 < b/\mathrm{km\,s}^{-1} < 100$. In § 2.1.5 (Figure 2.6) this reduces the initial dataset of 1264 at $z = 2$ to 845 absorbers that are effectively used for cutoff fitting.

As described in § 3.4.3, using the complete dataset results in a dramatic improvement in the precision compared to the measurements presented in § 2.4[7]. This improvement comes from the fact that the constraining power of the cutoff method depends only weakly on the number of absorbers in the b-$N_{\mathrm{H\,I}}$ distribution, as discussed in detail by Schaye et al. (1999) (see their Figure 14), and hence its precision does not scale as $\sqrt{N}$ as one would naively expect. In contrast, the advantage of the PKP method is that it delivers a precision that scales approximately as $\sqrt{N}$, delivering higher precision for larger datasets.

For a more direct comparison one can calculate what uncertainties we would expect for a dataset of 845 absorbers, i.e. the exact number of absorbers effectively used for cutoff fitting. Under the assumption of $\sqrt{N}$ scaling, our representative uncertainties for a dataset of 200 absorbers, for example, $\sigma_{\log T_0} = 0.055$ and $\sigma_\gamma = 0.074$, become smaller by a factor of $\sqrt{845/200}$, i.e $\sigma_{\log T_0} = 0.027$ and $\sigma_\gamma = 0.036$. In this case our result would be around a factor of two in $\log T_0$ and a factor of nearly three in γ more precise than cutoff fitting for 845 absorbers.

Indeed, the main limitation in PKP precision, which we have already encountered for the current dataset, is the number of simulations required to generate a model grid dense enough to deliver the implied precision. However, we believe that this is a surmountable problem given currently available computational resources.

Finally, we note that another complication associated with the cutoff fitting method is that one has to adopt a value of the column density $N_{\mathrm{H\,I},0}$ that corresponds to the mean density in order to relate the minimal Doppler parameter at this density $b_0 = b_{min}(N_{\mathrm{H\,I},0})$ to T_0. With this new approach we circumvent this issue, as we are sensitive to the shape of the b-$N_{\mathrm{H\,I}}$ distribution at all column densities. Furthermore, we show in § 2.4.4 that cutoff fitting is sensitive to the details of the iterative cutoff fitting method (least-squares or mean-deviation minimization), which can lead to differences in the results. In contrast, the Bayesian likelihood (eqn. 3.8) that provides the underpinnings of PKP

---

[7] This comparison may seem unfair since we marginalized our results over different pressure smoothing scales λ_P in § 2, which we do not do in this chapter. Nevertheless, this marginalization did not significantly impact their measurement precision, i.e. their uncertainties in T_0 and γ are dominated by the statistical error on the cutoff parameters.

does not require that one make these somewhat arbitrary choices.

3.5. Discussion and Summary

In this chapter we introduced a new method for inferring thermal parameters from the b-N_{HI} distribution of Lyα forest absorbers in the IGM, the PKP method. In contrast to a large body of previous work focused on analyzing a small subset of lines to fit the lower cutoff of the b-N_{HI} distribution, our new approach utilizes all available data and exploits the parameter sensitivity encoded in the full shape of this distribution. We generated a large grid of simulations of the Lyα forest encompassing a range of different thermal parameter models, and we fit the resulting mock spectra with `VPFIT`, generating a large database of absorption lines for each model. Our new method applies KDE to sets of discrete absorption lines to generate model b-N_{HI} distribution PDFs and then uses a PCA decomposition to create an emulator for this distribution that can be evaluated at any location in thermal parameter space. Using this emulator, we introduced a Bayesian likelihood formalism enabling parameter inference via MCMC. We conducted a pilot study demonstrating the efficacy of this new approach in the limit of a two-dimensional T_0 and γ measurement, whereby real observational data at $z = 2$ were compared to a grid of hydrodynamic simulations. The primary results of this chapter are as follows:

1. Using 160 mock measurements, we demonstrated that our statistical inference procedure delivers unbiased estimates of thermal parameters and reports valid uncertainties.

2. Our new method was applied to real observational data to measure the parameters of the TDR at $z = 2$. We found $\log T_0 = 4.092^{+0.050}_{-0.055}$ and $\gamma = 1.49^{+0.073}_{-0.074}$ using just a subset of 200 absorbers from the dataset presented in § 2, which roughly corresponds, in terms of path length, to a single Lyα forest spectrum at $z \simeq 2$.

3. For current dataset sizes at z=2, the PKP method can already deliver a precision on $\log T_0$ (γ) nearly two (three) times higher than the cutoff fitting method.

4. Thermal State of the IGM at $z = 0.1$

> *I always live in the present. I don't know the future and no longer have the past. The former oppresses me as the possibility of everything, the latter as the reality of nothing.*
>
> - Bernardo Soares, *Livro do Desassossego* 1930

In chapter 1, we discussed that the evolution of the thermal state of the IGM at low redshift is relatively simple, as no further reionization processes are expected to affect it. The thermal state of the intergalactic gas is dominated by cooling due to the universe's expansion and heating by photoionization. In this scenario the temperature at mean density is expected to drop as the universe expands, while γ approaches its asymptotic value of 1.7 (see Figure 1.2 for the evolution of the parameters of the TDR toward low redshift).

In the previous chapters of this thesis, we analyzed the Lyα forest in spectra of QSOs at $2 \leq z \leq 3.4$. This redshift range allows us to observe the Lyα forest in the optical using ground-based telescopes. However, the Lyα transition at $z \lesssim 1.7$ is shifted into the UV, where the atmosphere's transmission cuts off at around 3300 Å. Therefore, observations of the Lyα forest at low redshift require substantially more costly spectroscopic observations from space. Currently, there are two instruments on HST that are capable of performing UV spectroscopy in space, namely the space telescope imaging spectrograph (STIS) and the cosmic origins spectrograph (COS). STIS is sensitive in the near UV that probes the Lyα forest at $z \simeq 1$. COS on the other hand is sensitive to the far UV, probing the Lyα forest at $z \lesssim 0.5$.

In this study we will measure the TDR parameters (T_0 and γ) as well as the H I photoionization rate Γ_{HI} at $z = 0.1$ using publicly avilable HST/COS data[1], with the goal of testing theoretical predictions of the evolution of the thermal state of the IGM. Additionally, a measurement of Γ_{HI} at these redshift will reveal if we also observe a

---

[1] As no space based UV spectroscopy instrument is planned for the foreseeable future, these data might represent the last spectroscopic observations of the Lyα forest at low redshift in the near future. Note however, that there are missions planned such as the large UV optical infrared surveyor (LUVOIR) which is expected to launch sometime in the 2030's and Spektr-UV planned to launch in the middle of the 2020's that could fill this observational gap.

much higher $\Gamma_{\rm H\,I}$ than predicted by models. This effect has been observed by Kollmeier et al. (2014) based on the distribution of column densities of absorbers in the low redshift Lyα forest and was dubbed the “photon underproduction crisis”.

Here we apply a similar approach for inference as in Chapter 3, i.e. comparing the b-$N_{\rm H\,I}$ distributions of data and models by modeling the full distribution, but using a different estimator for the conditional probability density estimation. A new low redshift analysis of the Lyα forest is an interesting exercise to carry out, because the b-$N_{\rm H\,I}$ distribution has been previously used to study the effect of galactic feedback on the IGM (Viel et al. 2017; Gaikwad et al. 2017), but never in the context of measuring the parameters of the b-$N_{\rm H\,I}$ distribution.

This chapter is structured as follows. We introduce our dataset in § 4.1. Our hydrodynamic simulations, thermal parameter grid, and forward modeling approach are described in § 4.2. We describe our method for generating conditional probability distributions for our thermal grid in § 4.3. We perform the first low redshift measurements of T_0, γ and $\Gamma_{\rm H\,I}$ using the b-$N_{\rm H\,I}$ distribution in § 4.4. We summarize the results of this chapter in § 4.5.

4.1. Observational Data

Our observational dataset is composed of the publicly available version[2] of the high-quality medium resolution HST/COS survey presented in Danforth et al. (2016). Consisting of 82 spectra observed between 2009 and 2013 using the G130M and G160M gratings, this dataset represents the largest low redshift UV survey of the Lyα forest to date. The nominal resolution of COS is $R \sim 15000 - 20000$, which corresponds to roughly $15\,\mathrm{km\,s^{-1}}$, and has a non-Gaussian line spread function (LSF). Individual spectra were co-added taking into account both gratings in case they were available. In Danforth et al. (2016) these data were continuum fitted and absorption and emission lines were identified.

Although we will identify our own set of Lyα absorption profiles in § 4.1.1, we will mask out all intervening metal absorbers found by Danforth et al. (2016), which include both intervening and $z = 0$ MW absorbers.

We mask all emission lines, whereas the mask is adjusted by eye to include the full emission profile, and gaps in the wavelength coverage. An example of one spectrum and the metal masking procedure is shown in Figure 4.2. The continuum normalized spectrum is plotted in gray. The spectrum was fitted using `VPFIT` as will be described in § 4.1.1. The parts of the fitted spectrum shown in orange illustrate the segments that were masked based on the Danforth et al. (2016) line identifications. The unmasked fitted spectrum is shown in blue and consists of identified Lyα absorbers. To generate a

---

[2] http://archive.stsci.edu/prepds/igm/

Table 4.1. COS sightlines

ID	z_{qso}	SNR/pix.	LP	ID	z_{em}	SNR	LP
PG1011	0.0583	21.0	1	HE0056[a]	0.1641	17.0	1
RBS144	0.0628	15.0	1	H2356[a]	0.1651	10.0	2
PG1229	0.063	11.0	1	PG1048[a]	0.1671	13.0	1
MRK1513	0.063	20.0	1	1SJ1032[a]	0.1731	6.0	1
MR2251	0.064	26.0	1	PG2349[a]	0.1737	13.0	1
RXJ0503	0.064	11.0	1	PG1116[a]	0.1763	27.0	1
PG0844	0.064	13.0	1	PG1309[a]	0.1829	7.0	1
PKS2005	0.071	16.0	1	P1103[a]	0.186	10.0	1
TON1187	0.0789	12.0	1	PHL1811[a]	0.192	26.0	1
MRK478	0.0791	14.0	1	PHL2525[a]	0.199	11.0	2
PG0804	0.1	44.0	1	RBS1892[a]	0.2	12.0	2
I22456	0.1	35.0	1	PG1121[a]	0.225	11.0	1
UKS0242	0.1018	10.0	1	S50716[a]	0.2315	24.0	1
RBS542	0.104	52.0	1	PG0953[a]	0.2341	26.0	1
TONS210[a]	0.116	27.0	1	S092909[a]	0.24	10.0	1
PKS2155[a]	0.1165	36.0	2	RXJ0439[a]	0.243	11.0	1
Q1230[a]	0.117	41.0	1	F1010[a]	0.2558	10.0	1
MRK106[a]	0.123	17.0	1	PKS1302[a]	0.2784	18.0	1
PG1435[a]	0.126	9.0	1	TON580[a]	0.2902	13.0	1
I06229[a]	0.129	19.0	1	H1821[a]	0.2968	36.0	1
MRK876[a]	0.129	45.0	1	S09255B[a]	0.3295	10.0	1
PG0838[a]	0.131	18.0	1	PG1001[a]	0.3297	14.0	1
PG1626[a]	0.133	18.0	1	PG0832[a]	0.3298	10.0	1
Q0045[a]	0.134	21.0	1	PG1216[a]	0.3313	14.0	1
PKS0558[a]	0.1372	10.0	1	3C66A[a]	0.3347	12.0	2
PG0026[a]	0.142	10.0	1	RXJ2154[a]	0.344	15.0	1
S135712[a]	0.15	9.0	1	B0117[a]	0.3489	16.0	1
PG1115[a]	0.1546	14.0	1	PG1049[a]	0.3599	8.0	1
RXJ0956[a]	0.155	11.0	1	1ES1028[a]	0.3604	10.0	1
PG1307[a]	0.155	12.0	1	S094952[a]	0.3656	7.0	1
3C273[a]	0.1583	65.0	1	1ES1553[a]	0.414	17.0	1
PG0157[a]	0.1631	11.0	1	HE0435[a]	0.43	5.0	1

[a] These sightlines contribute to the pathlength $0.06 < z < 0.16$.

Table 4.2. COS sightlines

ID	z_{qso}	SNR/pix.	LP	ID	z_{em}	SNR	LP
PG1222[a]	0.432	13.0	2	PMN2345	0.621	7.0	2
TON236[a]	0.45	8.0	1	HE0238	0.631	11.0	1
PG0003[a]	0.4509	11.0	1	3C263	0.646	12.0	1
HE0153[a]	0.451	11.0	1	PKS0637	0.65	10.0	1
PG1259	0.4778	14.0	1	S080908	0.6563	6.0	1
HE0226	0.4934	14.0	1	3C57	0.6705	6.0	1
HS1102	0.5088	7.0	1	PKS0552[b]	0.68	0	1
PKS0405	0.574	19.0	1	SBS1108[b]	0.7666	0	1
PG1424	0.6035	11.0	1	SBS1122[b]	0.852	0	1

[a] These sightlines contribute to the pathlength $0.06 < z < 0.16$.

[b] These sightlines have very small coverage of the Lyα forest, as the COS instrument cuts off at 1800 Å.

mask, a window was adjusted by eye in order to remove each metal absorber based on its central wavelength positions. The species of each metal absorber is written in the corresponding position. This masking is done in post-processing, which means that we first apply `VPFIT` to the spectra and then remove the absorbers that fall within the masked regions. In contrast to high redshift ($z > 1$) spectra, the low redshift Lyα forest consists of mostly unabsorbed continuum and few, mostly isolated, absorption lines.

In Figure 4.1 we show the redshift path that is covered by the data (after all rejections). The lines are color coded based on the sightline SNR and gaps in the spectra correspond to masked regions. It is noticeable that some gaps appear at the same wavelength for different sightlines. These are due to metal absorbers in the MW that fall always at $z \sim 0$, contrary to intervening metals that could be at any redshift between the MW and the QSO.

4.1.1. Voigt-Profile Fitting

We used our fully automated `VPFIT` wrapper described in § 2.1.2 for fitting absorption profiles to the COS dataset described above. This time we applied the current `VPFIT` version 11.1[3]. As described in § 2.1.2 our wrapper routine controls `VPFIT` with the help of the `VPFIT` front-end/back-end programs RDGEN and AUTOVP`IN`. Each absorption line is parametrized as follows: absorption redshift $z_{\rm abs}$, Doppler parameter b, and column density $N_{\rm HI}$ corresponding to the Lyα transition. We set up `VPFIT` to explore the range of parameters $1 \leq b/{\rm km\,s^{-1}} \leq 300$ and $11.5 \leq \log(N_{\rm HI}/{\rm cm^{-2}}) \leq 18$. `VPFIT` will vary these parameters and potentially fit for additional lines until the χ^2 with respect to the data is minimized. For more details on the fitting procedure refer to section 2.1.2.

One important difference with respect to the UVES/HIRES spectra fitted in Chapters 2 and 3 is that the LSF of the COS instrument has a moderate resolution of $15{\rm km\,s^{-1}}$ (compared to a resolution of $6{\rm km\,s^{-1}}$ in the UVES/HIRES dataset). The COS resolution is of the order of the typical Doppler broadening parameter in the Lyα forest. The COS LSF is also non-Gaussian in shape and exhibits significant wings which result in pixel-to-pixel correlations across larger scales. Therefore, it is particularly important to take into account the instrument's LSF, otherwise one might be overestimating the width of absorbers due to LSF effects. We do this by splitting all spectra at observed $\lambda = 1460\,$Å and retrieving the instrument's LSF for both of COS arms separately at 1300 Å for the G130M grating and 1600 Å for the G160M grating with the help of the `linetools` package[4]. The effective resolution will also be dependent on the lifetime position of the observation, which we take into account when running `VPFIT`. For

---

[3] VPFIT: http://www.ast.cam.ac.uk/~rfc/vpfit.html

[4] https://github.com/linetools/linetools

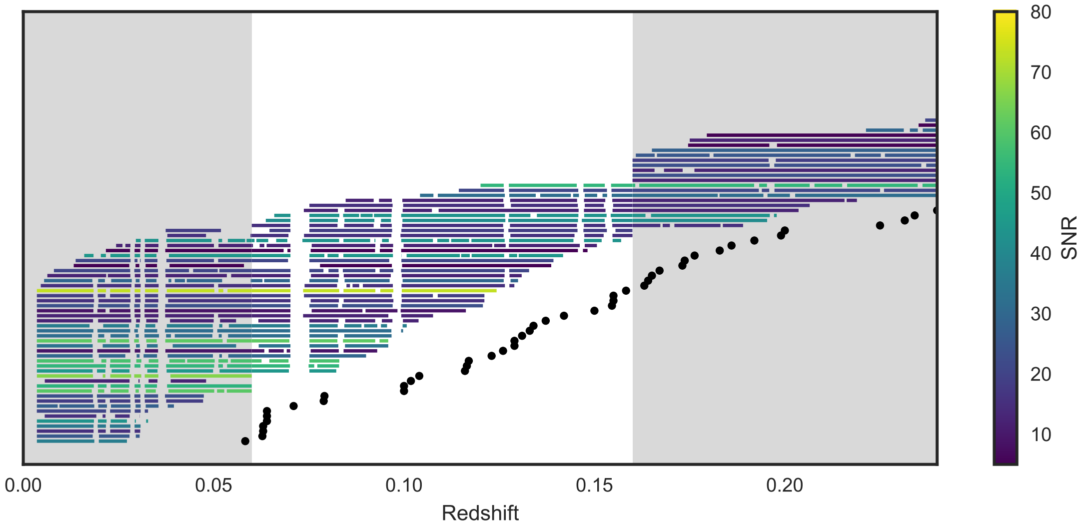


Figure 4.1. COS Lyα forest sightlines used in this chapter. The sample is described in § 4.1. The corresponding QSO redshifts are marked as black points. Regions that were excluded are shown as gaps. The white area indicates the bin $0.06 < z < 0.16$ that will be used in this analysis. Gray regions are not used.

simplicity we assume that the LSF does not vary as a function of wavelength. For the spectra analyzed in this chapter, only the G130M grating resolution is used, as the Lyα forest at $z = 0.1$ lies completely in this region.

After fitting the entirety of the COS dataset, we select all the absorbers in the range $0.06 < z_{abs} < 0.16$ for further analysis of the thermal state of the IGM at $z = 0.1$. An example of the VP-fitted spectrum of the COS sightline PHL1811 is shown in Figure 4.1 in blue (segments of the spectra previously flagged as metal absorption are marked in orange). The b-$N_{\rm H\,I}$ distribution of all absorbers in this redshift range is shown in Figure 4.6.

4.2. Simulations

In order to generate a set of model b-$N_{\rm H\,I}$ distributions for carrying out a measurement of the thermal paramters at $z = 0.1$, we use part of the publicly available THERMAL suite of Nyx simulations (Almgren et al. 2013; Lukić et al. 2015) as introduced in § 2.2 and 3.1. Because of peculiarities in the COS LSF we perform a slightly different forward modeling approach.

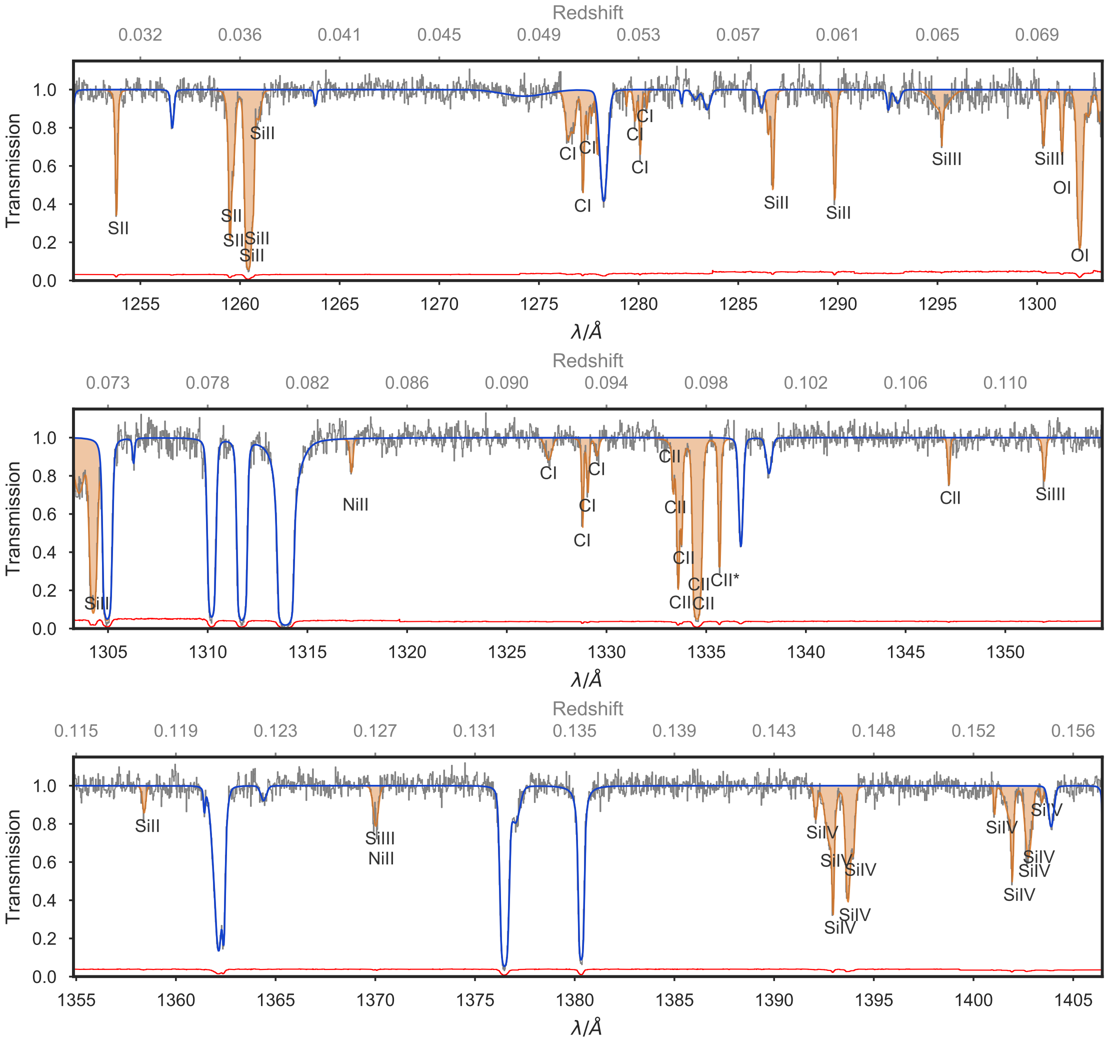


Figure 4.2. Illutration of the COS spectrum PHL1811. Each panel depicts a segment of the QSO spectrum. The original spectrum is shown in gray, while a model spectrum based on line fitting (described in § 4.1.1) is shown in blue (masked spectrum) and orange (only masked regions of the spectrum). The line IDs used to generate the mask are from Danforth et al. (2016) and the metal species is denoted at the absorber wavelength.

4.2.1. Skewer Generation

Our previous approaches for generating skewers was based on rescaling the optical depth field of our simulations to match a given mean transmission, which corresponds to rescaling the total ionization rate by a factor[5]. This is acceptable at high redshift, as the gas is predominantly photoionized. At low redshift one expects a substantial increase in the abundance of collisionally heated gas. Rescaling the optical depth would effectively result in a rescaling of the collisional ionization rate by the same factor.

Therefore, in order to generate skewers at these redshifts, we have to apply a slightly different recipe. Note that we cannot treat the amount of collisional ionization as a free parameter, as it is strongly dependent on the hydrodynamic properties of the gas. We generate skewers by recomputing the optical-depth field τ through the simulations assuming different values of $\Gamma_{\mathrm{H\,I}}$ and keeping collisional ionization fixed.

4.2.2. Thermal Parameter Grid

We construct a thermal grid consisting of a subset of 48 simulation snapshots at $z = 0.1$ with different combinations of T_0 (spanning $3.3 \lesssim \log T_0 \lesssim 4.1$) and γ (spanning $1.2 \lesssim \gamma \lesssim 2.4$). This grid is shown in the left panel of Figure 4.3 and its irregular shape simply reflects that the thermal parameters result from different thermal history inputs in the simulations. For each of these models we generate skewers using 7 different values of the H I photoionization rate Γ_{HI} (from $\Gamma_{\mathrm{HI}} = 10^{-13.383}\,\mathrm{s}^{-}$ to $10^{-12.932}\,\mathrm{s}^{-}$ in logarithmic steps of 0.075 dex, see right panel of Figure 4.3). In total our 3D grid consists of 336 models.

Any of the effects of reionization are expected to fade out at redshifts $z < 0.4$, which results in λ_P becoming fully degenerate with T_0 and γ. Note that in order to add the pressure smoothing scale λ_P as a parameter in the previous chapters, we used the technique described in Kulkarni et al. (2015) that consists of fitting the cutoff in the power spectrum of the real-space Lyα flux F_{real} (see details in § 2.2). These have not yet been computed for our current grid, so this first exploration of the low redshift IGM will include T_0, γ and Γ_{HI}.

4.2.3. Forward Modeling of Noise and Resolution

Before convolving the skewers with the LSF, we chose a random COS spectrum and stitched as many simulated lines of sight as necessary to fill the pathlength of chosen observed QSO. Next, this stitched together spectrum is binned to the wavelength scale of the chosen COS sightline.

As discussed in § 4.1, the COS instrument has a particular LSF shape. Therefore, we have to account for the peculiarities of the COS resolution in our forward models in order to be able to compare the models with the data. We modeled the wavelength

---

[5] This is true only in the optically thin regime where the Voigt profile can be well approximated by a Gaussian, but most of the Lyα forest absorbers fulfill this criterion.

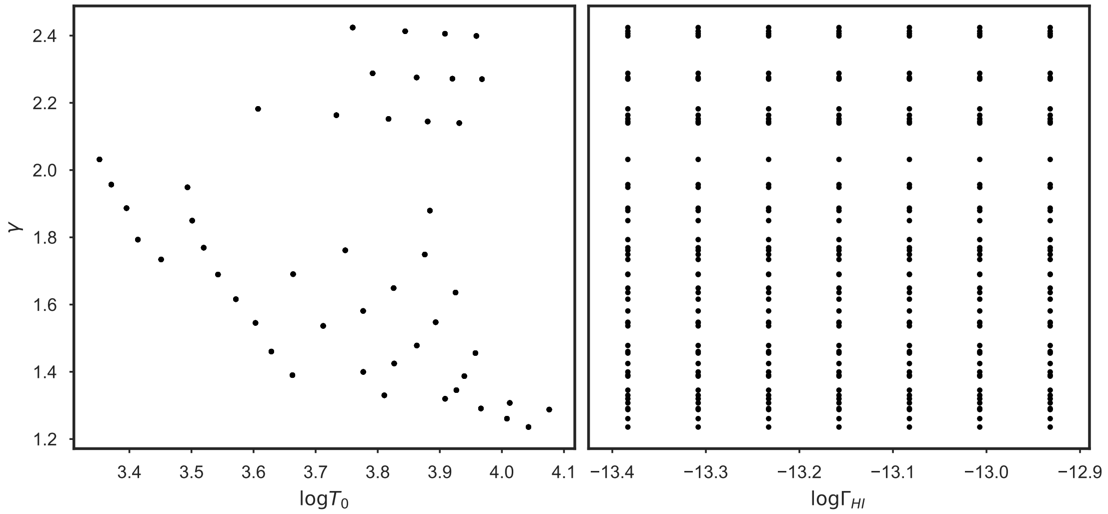


Figure 4.3. Thermal grid from snapshots of hydrodynamic simulations from the THERMAL suite at $z = 0.1$ used in this low redshift study.

dependency of the COS LSF by modeling each output pixel as a convolution between the input skewer and the LSF for this wavelength as obtained from `linetools`, taking into account grating and lifetime position. The LSF is tabulated for up to 160 pixels in each direction and was interpolated to the pixel scale of the simulation.

We propagate the noise properties of the data to our simulated spectra on a pixel-by-pixel basis by sampling from a Gaussian with $\sigma = \epsilon_i$, with ϵ_i being the data noise vector value at the current pixel.

For each simulation snapshot we computed 6000 of these stitched together simulated skewers in order to generate a well populated b-N_{HI} distribution.

4.3. Method for Emulating the Full b-N_{HI} Distribution

In order to model the b-N_{HI} distribution as a whole, we follow the same approach as in Chapter 3 and estimate $P(\log N_{\mathrm{HI}}\,, \log b \mid \log T_0, \gamma)$ from the `VPFIT` output of forward modeled spectra from hydrodynamic simulations. The main difference is in the method used for density estimation. While in Chapter 3 we estimated and interpolated PDFs using a combination of KDE, PCA and GP, in this chapter we will use the density-estimation likelihood-free inference (DELFI) method (Alsing et al. 2018, 2019) for this task.

DELFI uses neural density estimation (NDE), based on a masked autoregressive flow technique (Papamakarios et al. 2017), to learn the sampling distribution of the data as

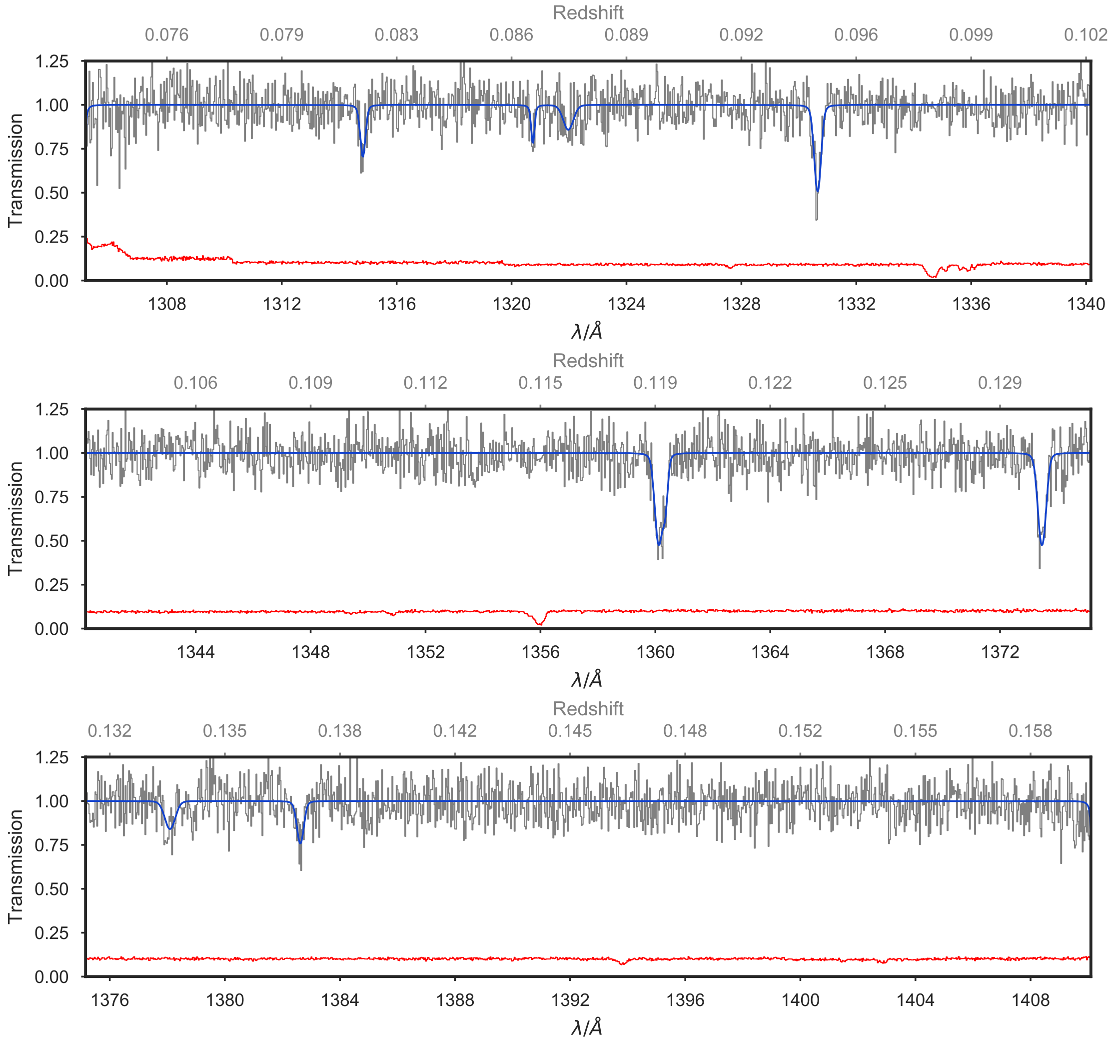


Figure 4.4. Illustration of one of our forward modeled spectra. Each panel depicts a segment of the spectrum. The original spectrum is shown in gray, while a model spectrum based on line fitting (described in § 4.1.1) is shown in blue. Our simulated spectra are not masked, given that metal absorption is not modeled. This particular model was forward modeled in order to reproduce the noise properties of the COS sightline RXJ0439.

a function of parameters $P(d|\theta)$, with θ being $\log T_0$, γ and $\log \Gamma_{\mathrm{HI}}$ and data d being $\log N_{\mathrm{HI}}$ and $\log b$ in our thermal grid[6].

---

[6] DELFI also has the option to apply different data compression and active learning methods to optimize the data and parameter space sampling which we do not use as we have pre-chosen our summary

As a sanity check, we compare the 2D inference carried out in § 3.4 to a new measurement performed using the same models and data, but with DELFI as an emulator. The comparison of the resulting contours are shown in Figure 4.5. Using DELFI we measure $\log T_0 = 4.104^{+0.031}_{-0.033}$ and $\gamma = 1.446^{+0.043}_{-0.044}$, whereby the uncertainties are calculated based on the 16th and 84th percentiles of the marginalized distributions. Both density estimation and emulation approaches deliver consistent results, with DELFI having smaller statistical uncertainties. This is presumably because DELFI does not apply a global smoothing for every point in the distribution and therefore delivers more precise unsmoothed PDFs.

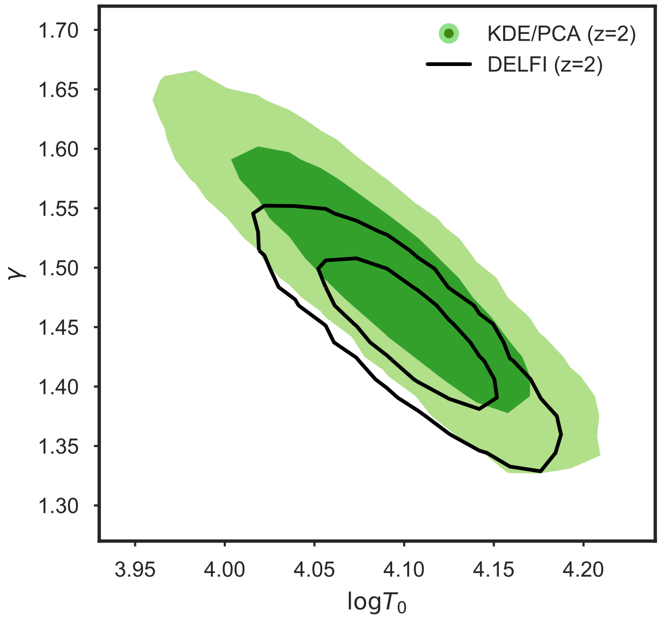


Figure 4.5. Comparison of the MCMC posterior of measurements using the same observational data at $z = 2$ (§ 3.4), but different methods for emulating the b-N_{HI} distribution models. The 1σ and 2σ contours are shown for the measurement using PKP (green) and DELFI (black lines) as an emulator. We achieve better constrains using DELFI, presumably because it does not require a global smoothing length and can therefore model $P(\log N_{\mathrm{HI}}\,, \log b \mid \log T_0, \gamma)$ more precisely.

---

statistics (the b-N_{HI} distribution) at a fix grid of thermal parameters.

4.4. A Measurement of Thermal Parameters at z=0.1

4.4.1. Likelihood

In this study we expand upon our likelihood definition from § 3.2.4 and additionally model the number of absorbers as an additional piece of data using a model for the Poisson rate of absorbers in the simulations.

Assuming that the PDFs are normalized such that

$$\iint P(b, \mathrm{N_{HI}})\, \mathrm{dN_{HI}}\, \mathrm{d}b = 1, \tag{4.1}$$

the likelihood of observing one or zero absorbers in an infinitesimally small grid in b and $\mathrm{N_{HI}}$ with $n + m$ number of cells, whereas n is the number of cells populated by one absorber and m is the number of unpopulated cells, under the assumption of a Poisson distribution, can be written as

$$\begin{aligned} \mathcal{L} =& P(\mathrm{data}|\mathrm{model}) \qquad (4.2)\\ =& \left(\prod_{i=1}^{n} \mu_i \, e^{-\mu_i}\right)\left(\prod_{j=n+1}^{m+n} e^{-\mu_j}\right). \end{aligned}$$

The rate μ_i of populating a cell is defined as:

$$\mu_i = \frac{\mathrm{N_{model}}}{\Delta z_{\mathrm{model}}} P(b_i, \mathrm{N_{HI}}_i)\, \Delta \mathrm{N_{HI}}_i\, \Delta b_i\, \Delta z_{\mathrm{data}}, \tag{4.3}$$

where the $\mathrm{N_{model}}$, $\Delta z_{\mathrm{model}}$ and Δz_{data} are the total number of lines (sets of b and N_{HI} values), the total redshift path in the simulations and the total redshift path covered by the data, respectively.

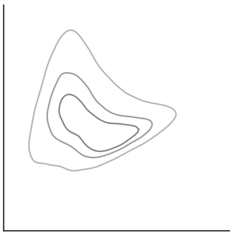
Using the logarithm, we can reduce eqn. 4.2 to

$$\begin{aligned} \ln \mathcal{L} &= \sum_{i=1}^{n} (\ln(\mu_i) - \mu_i) - \sum_{j=n+1}^{m+n} \mu_j \qquad (4.4)\\ &= \sum_{i=1}^{n} \ln(\mu_i) - \sum_{k=1}^{m+n} \mu_k \end{aligned}$$

$$(4.5)$$

The sum over k is simply an integral over eqn. 4.3, and since the integral over $P(b, \mathrm{N_{HI}})$ $\mathrm{dN_{HI}}$ db is unity (eqn. 4.1), we finally derive

$$\ln \mathcal{L} = \sum_{i=1}^{n} \ln(\mu_i) - \frac{\mathrm{N_{model}}}{\Delta z_{\mathrm{model}}} \Delta z_{\mathrm{data}}, \tag{4.6}$$



which is the likelihood formalism that will be used in this chapter.

In order to use the above likelihood, we will limit the simulation and data output to the region $12.5 \leq \log N_{\rm HI} \leq 14.5$[7] and $0.5 \leq \log b \leq 2.3$. This ensures that the quantities $\Delta \mathrm{N}_{\mathrm{HI}i}$ and Δb_i are not model dependent and can be safely ignored in the inference procedure. Furthermore it ensures that eqn. 4.1 is satisfied for all models.

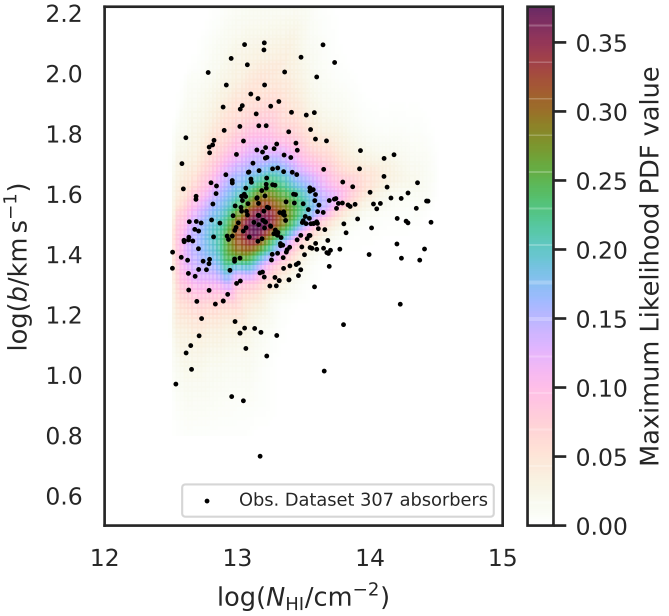


Figure 4.6. The b-$N_{\rm HI}$ distribution from the COS dataset at $0.06 < z < 0.16$ (black circles). An emulated $P(\log N_{\rm HI}\ , \log b \,|\, \log T_0, \gamma, \log \Gamma_{\rm HI})$ based on the median values of the marginal distributions of the corresponding MCMC posterior ($\log T_0 = 3.628$, $\gamma = 1.990$ and $\log \Gamma_{\rm HI} = -13.115$, see Figure 4.7) is shown for comparison.

The total redshift pathlength of the dataset $\Delta z_{\rm data}$ was calculated from the COS dataset at $0.06 < z < 0.16$ and is $z_{\rm data} = 2.058$. For each model in our thermal grid we calculated the total redshift pathlength $\Delta z_{\rm model}$ to be used for Voigt profile fitting, as well as the total number of absorbers $N_{\rm model}$. During inference, we emulate the quantity $N_{\rm model}/\Delta z_{\rm model}$ using a similar Gaussian process interpolation scheme as described in § 3.2.3. We constructed the emulator using smoothing lengths of 40% of our thermal grid length in each dimension and used a white-noise term of 10^{-3}.

---

[7] As discussed in § 2.1.3 this region in $N_{\rm HI}$ is sensible, because it avoids barely detected lines at low $N_{\rm HI}$, as well as saturated lines and LLS at high $N_{\rm HI}$ that are not modeled in the simulations, but are present in the data.

4.4.2. Parameter Estimation

We perform inference by sampling the parameter space via MCMC using the python package `emcee` (Foreman-Mackey et al. 2013) which implements the affine-invariant sampling technique (Goodman & Weare 2010) to sample the posterior probability distribution. As discussed in 4.2, we do not treat λ_P as a free parameter in this first exploration of the low redshift data. Furthermore, we assume uniform (flat) priors for $\log T_0$, γ and $\log \Gamma_{\mathrm{HI}}$. The algorithm is tied to the likelihood definition from eqn. 4.4 which takes into account the number of absorbers per pathlength in the data and simulations. This means that the $N_{\mathrm{model}}/\Delta z_{\mathrm{model}}$ GP emulator is evaluated at the current walker positions during the MCMC sampling.

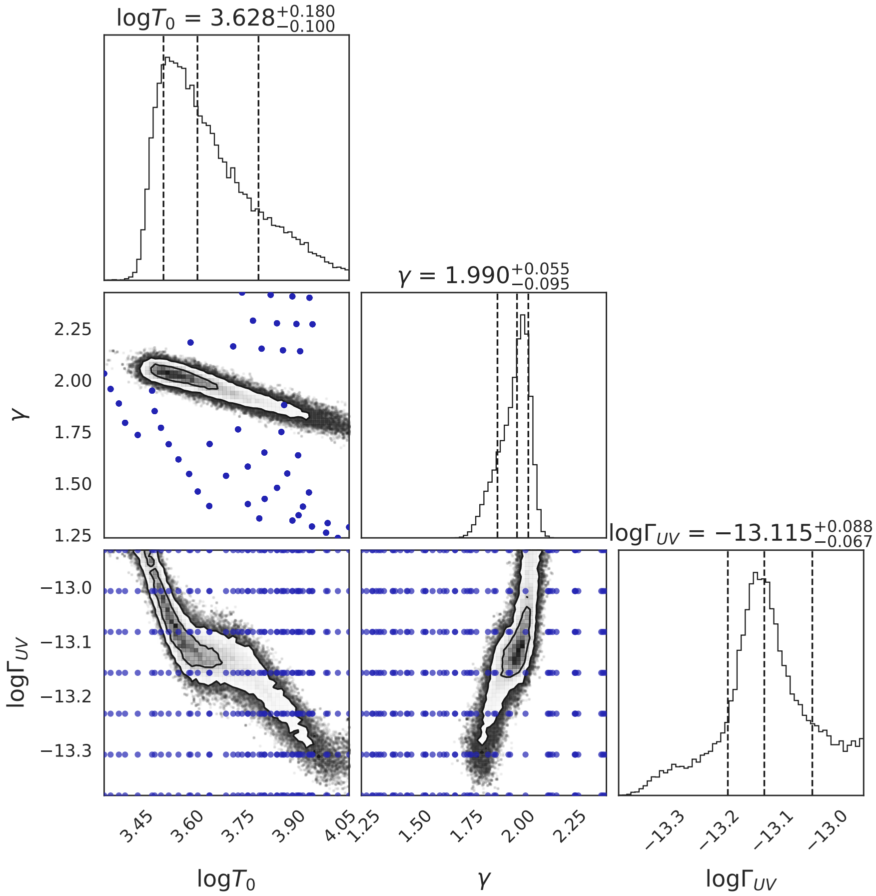


Figure 4.7. MCMC posterior for the fit of the b-N_{HI} distribution from the COS dataset at $0.06 < z < 0.16$ (absorbers shown in Figure 4.6) based on models of $P(\log N_{\mathrm{HI}}, \log b \mid \log T_0, \gamma, \log \Gamma_{\mathrm{HI}})$ estimated using DELFI (see § 4.3). Projections of the thermal grid used for generating models are shown as blue circles.

The b-$N_{\mathrm{H\,I}}$ distribution retrieved for the observational data is shown in Figure 4.6 as black points. We illustrate the model $P(\log N_{\mathrm{H\,I}}\,, \log b \,|\, \log T_0, \gamma, \log \Gamma_{\mathrm{H\,I}})$ estimated by DELFI at the position of the median values of the MCMC posterior as a color-coded map. The MCMC posterior distributions are presented in Figure 4.7.

We obtain $\log T_0 = 3.628^{+0.180}_{-0.100}$, $\gamma = 1.990^{+0.055}_{-0.095}$ and $\log \Gamma_{\mathrm{H\,I}} = -13.115^{+0.088}_{-0.067}$ from the marginalized distributions. In Figure 4.8 we illustrate how our measurements com-

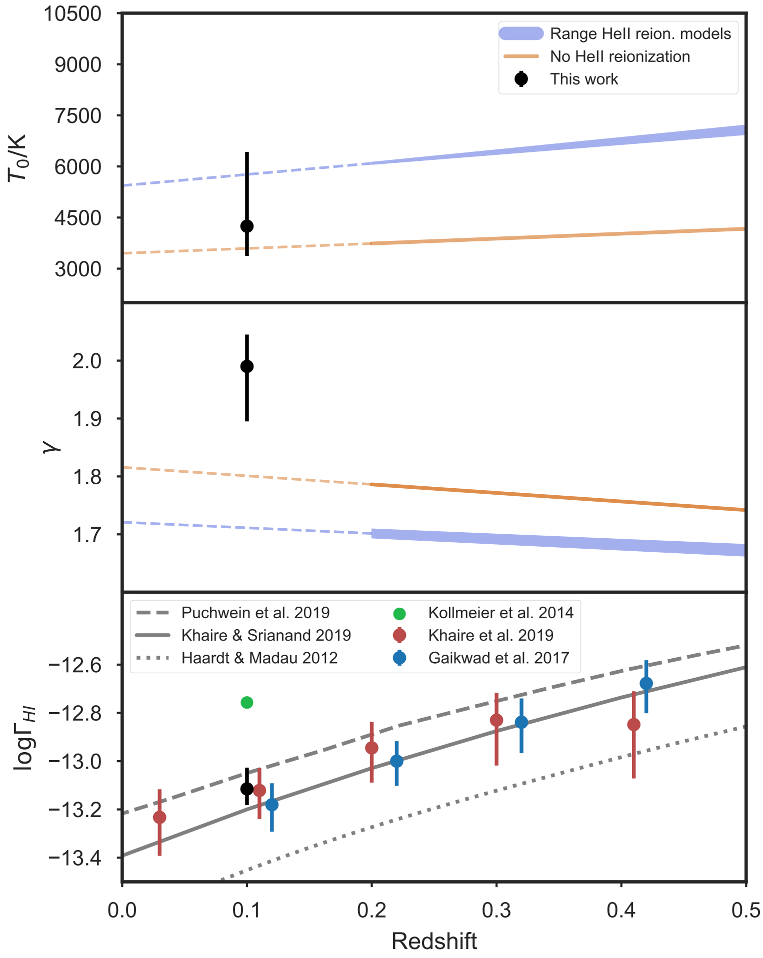


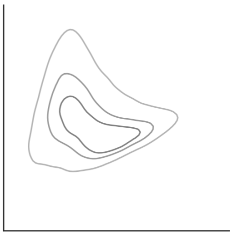
Figure 4.8. Comparison of our measurements (black circles) with models, simulations and other measurements from the literature. In the upper and middle panels we plot a region with the maximum and minimum T_0 and γ values in the thermal histories of all Nyx runs that include He II reionization (blue, the dashed line is an extrapolation of the evolution) as well as an evolution where no He II reionization happens (orange, the dashed line is an extrapolation of the evolution). We compare our $\Gamma_{\mathrm{H\,I}}$ to a variety of models and measurements in the bottom panel.

pare with model expectations and measurements in the literature. The measurements carried out in this work are shown as black circles. In the upper and middle panels we plot a region with the maximum and minimum T_0 and γ values in the thermal histories of all Nyx runs that include He II reionization (blue) as well as an evolution where no He II reionization happens (orange). These simulations were run down to $z = 0.2$. For illustration, an extrapolation of the evolution is plotted as a dashed line in both cases. The lower panel shows the evolution of $\Gamma_{\mathrm{H\,I}}$ as a function of redshift. The predictions by Puchwein et al. (2019) (dashed gray), Khaire & Srianand (2019) (solid gray), and Haardt & Madau (2012) (dotted gray) are also shown, as well as the $\Gamma_{\mathrm{H\,I}}$ measured by Khaire et al. (2019) (red, using the power spectrum), Kollmeier et al. (2014) (green, using the distribution of $N_{\mathrm{H\,I}}$), and Gaikwad et al. (2017) (blue, using the power spectrum).

Our measurement of T_0 is in general agreement with expectations that the IGM at low redshift should have cooled down to a temperature at mean density below ~10000 K due to cosmic expansion. The current precision, given the size of our dataset, indicates that the measured T_0 does not distinguish between the scenarios of He II reionization and no He II reionization. Our measurement of γ is significantly higher than theoretical expectations, which indicates a higher temperature contrast between overdensities and underdensities in the IGM. Our measurement of $\Gamma_{\mathrm{H\,I}}$ is consistent with the prediction by Puchwein et al. (2019) and slightly lower than the prediction by Khaire & Srianand (2019). When comparing our measurement to Kollmeier et al. (2014), we do not see evidence for a photon underproduction crisis in UVB models at $z = 0.1$.

4.5. Discussion and Summary

In this chapter, I performed a measurement of the thermal state of the IGM for the first time using the b-$N_{\mathrm{H\,I}}$ distribution at low redshift. This was achieved by applying the new method described in Chapter 4 that is sensitive to the full shape of the b-$N_{\mathrm{H\,I}}$ distribution. We measure a value of T_0 that is in agreement with theoretical expectations but does not clearly distinguish between the scenarios of where He II reionization took place or not. The value of γ is significantly higher than expected. In this analysis I also included the hydrogen photoionization rate $\Gamma_{\mathrm{H\,I}}$ as a free parameter and recovered a value that is in good agreement with current predictions. Our measurement of $\Gamma_{\mathrm{H\,I}}$ disfavors claims of a photon underproduction crisis in current models of the UVB.



5. Conclusions and Outlook

And in the end
The love you take
Is equal to the love you make

- Lennon/McCartney, *The End* 1969

In this thesis I carried out measurements of T_0 and γ, the parameters that describe the temperature-density relation (TDR) of the intergalactic gas (introduced in Chapter 1), as a function of redshift. These parameters were retrieved by carefully studying the distribution of Doppler parameters b and column densities N_{HI} (b-N_{HI} distribution) of absorbers in the Lyα forest. The approach I followed consists of comparing observational data to the b and N_{HI} output of state of the art hydrodynamic simulations. The methodology of each chapter can be summarized as follows:

- **In Chapter** 2 I presented a new version of a classic approach of measuring thermal parameters from the b-N_{HI} distribution, namely by quantifying the position of the lower cutoff in the distribution. I expanded on previous work by significantly increasing the number of observed sightlines used (75 UVES and HIRES spectra) and by carefully calibrating the relationship between cutoff and thermal parameters with the help of hydrodynamic simulations. Furthermore, consideration is given to the effect of the pressure smoothing scale λ_P on the measurement uncertainties. With this method I constrained the evolution of T_0 and γ in the redshift range $2 \le z \le 3.4$.
- **In Chapter** 3 I developed, tested and applied a new method for carrying out measurements of thermal parameters using the b-N_{HI} distribution that goes beyond the cutoff and instead treats this distribution as a whole. This approach is more accurate when compared with cutoff fitting and delivers measurements with significantly smaller statistical uncertainties for current dataset sizes. This method was applied to observational data at $z = 2$.
- **In Chapter** 4 I performed a measurement of the thermal state of the IGM at $z = 0.1$ for the first time using the b-N_{HI} distribution by applying the approach of Chapter 3 to spectra from the HST/COS instrument.

The individual measurements of the parameters of the TDR that were obtained in each chapter are summarized in Figure 5.1. I compare these to the range of minimum/maximum values of γ and T_0 allowed by different thermal histories from the Nyx hydrodynamic simulation runs that include He II reionization (blue area). For comparison, I show the simulated thermal evolution without He II reionization in orange. The measurements from Chapter 2 (blue, using the cutoff fitting technique) are strongly inconsistent with a scenario where He II reionization does not happen, due to significantly higher T_0. Furthermore, these measurements indicate that inverted TDR (with $\gamma < 1$) are disfavored. The result of Chapter 3 (red) indicates that, if one uses the whole shape of the b-N_{HI} distribution I measure a TDR at $z = 2$ that is consistent with the measurement from Chapter 2. Furthermore, one can achieve better constraints with this novel method, even when using a fraction of the available data. In Chapter 4 (black), I carried out measurements of the thermal state of the IGM for the first time using the b-N_{HI} distribution at low redshift. This was done by using the new method described in Chapter 3, combined with neural density estimators, that is sensitive to the full shape of the b-N_{HI} distribution. I recover a value of T_0 that is in agreement with theoretical expectations, but does not distinguish between the scenarios of He II reionization. The value of γ is significantly higher than expected, indicating a stronger

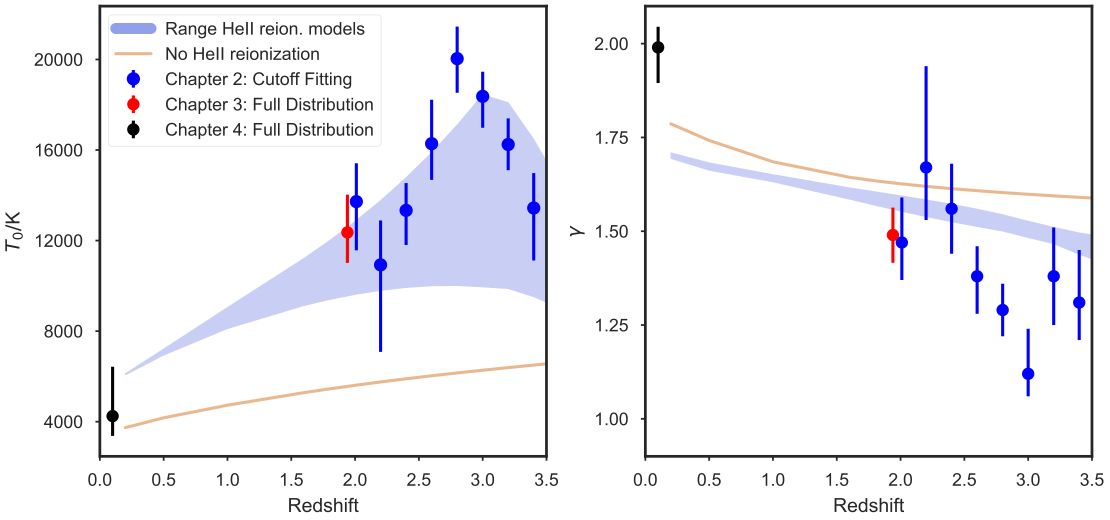


Figure 5.1. Summary of all measurements of the parameters of the TDR presented in this thesis. The measurements performed in each chapter are plotted as circles. For comparison, the range of minimum/maximum values of γ and T_0 allowed by different thermal histories from the Nyx hydrodynamic simulation runs that include He II reionization is shown in blue. Additionally, the simulated thermal evolution without He II reionization is shown in orange.

temperature contrast between overdensities and underdensities in the low redshift IGM. In this analysis I also included the hydrogen photoionization rate $\Gamma_{\mathrm{H\,I}}$ as a free parameter and recovered a value that is in good agreement with current predictions, disfavoring claims of a photon underproduction crisis in current UVB models.

This thesis demonstrates the statistical power of the b-$N_{\mathrm{H\,I}}$ distribution and shows that it is one of the most powerful tools for studying the evolution of the thermal state of the IGM at redshifts $z < 4$.

Outlook

In Chapter 3 I have presented a new method for constraining thermal parameters in the IGM based on the full shape of the b-$N_{\mathrm{H\,I}}$ distribution. This method can be expanded to include other parameters that affect the shape of this distribution (provided that these can be modeled in simulations). One interesting case would be to include a measurement of the pressure-smoothing scale λ_P, as only few studies were able to measure λ_P at $z > 1.8$ (Rorai et al. 2017b; Walther et al. 2019).

Furthermore, one could use the b-$N_{\mathrm{H\,I}}$ distribution in order to analyze IGM models with additional physics such as blazar heating (Puchwein et al. 2012; Sironi & Giannios 2014; Lamberts et al. 2015) or galaxy formation feedback (Sorini et al. 2018), as these are expected to affect the shape of absorption profiles in the Lyα forest.

This new methodology is readily applicable to the $2 < z < 4$ Lyα forest, as shown by the study at $z = 2$, as well as to existing Hubble Space Telescope Cosmic Origins Spectrograph (HST/COS) UV spectra (e.g. Danforth et al. 2013, 2016) that probe the Lyα forest at $z \lesssim 0.5$. In order to fill the redshift gap in the measurements presented here, this approach could also be applied on STIS data that probe the Lyα forest around $z = 1$.

The methods for conditional density estimation presented in this thesis can be applied to any set of observable summary statistics in simulations that change smoothly with the parameters of interest. This includes fields outside of cosmology where the comparison between forward modeled simulations and observed data are an important tool, such as galactic archeology, galactic dynamics or planetary disk formation.

Appendices

A. Cutoff Fitting

A.1. QSO Continuum Placement Effect on the Cutoff Measurement

The continua of the QSOs in our sample were placed based on the regions of the spectra that have no clear absorption and are therefore subject to uncertainty. A misplacement of the continua could certainly have an effect on the corresponding optical depth of a line (and therefore on the line-profile parameters). Different studies show that for high SNR and resolution data, the statistical uncertainty of the continuum placement is of the order of a few percent at $z < 4$ (Kirkman et al. 2005; Kim et al. 2007; Dall'Aglio et al. 2008; Faucher-Giguère et al. 2008b). We assume that our typical continuum uncertainty is of the order of $\sim 2\%$ / $\sim 5\%$ for $z < 3$ / $z > 3$ sightlines.

To address the effect of continuum misplacement in our study, we analytically estimate how a shift of 2% and 5% in the continuum affects the typical line in our sample. This is done by calculating the optical depth at line center (Meiksin 2009) using eqn. 1.16 for lines with different column densities and a typical width of $b = 19\,\mathrm{km\,s^{-1}}$, and converting it to flux at line center $F_{lc} = \exp(-\tau_{lc})$. This flux is shifted by 2% and 5% to mimic the effect of misplacement of the continuum; then, doing the reverse operations and keeping b fixed, we compute the corresponding $\log N_{\mathrm{H\,I}}$ values.

As illustrated in Figure A.1, for a continuum shift of 2% the corresponding shift in $\log N_{\mathrm{H\,I}}$ is generally smaller than the uncertainty in $\log N_{\mathrm{H\,I}}$ reported by `VPFIT` within our cutoff fitting range. For a continuum misplacement of 5%, the `VPFIT` uncertainty becomes comparable to the continuum misplacement effect at column densities $\log(N_{\mathrm{H\,I}}/\mathrm{cm^{-2}}) = 13$ and exceeds it at lower $N_{\mathrm{H\,I}}$.

Given the small effects on the column densities at lower redshift, the errors due to continuum placement can be neglected. At lower continuum placement precision ($z > 3$) this effect can influence the lower column densities, but our cutoff fitting algorithm is likely not very sensitive to this, because it is driven by absorbers with better constrained parameters at higher column densities.

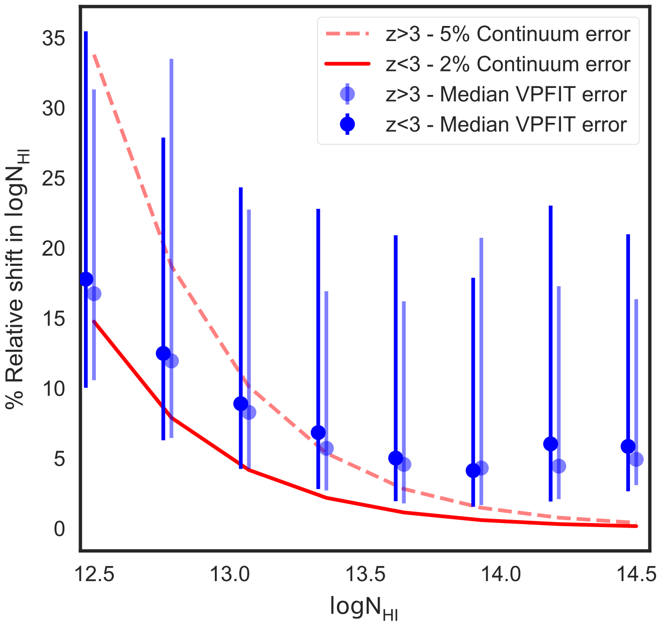


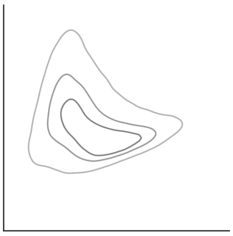
Figure A.1. Estimation of the relative shift in $N_{\rm HI}$ caused by a misplacement of the continuum for 2% (red solid) and 5% (red dashed) and the VPFIT error distributions for our list of absorbers with $z > 3$ (blue) and $z < 3$ (light-blue). With a continuum uncertainty of 2%, the effect is smaller than the typical error in $N_{\rm HI}$ reported by VPFIT. A continuum uncertainty of 5% affects column densities $\log N_{\rm HI} < 13$. Figure credit: Hiss et al. (2018).

A.2. Impact of Uncertainties in the Mean Flux

We describe in § 2.2 how our simulations are rescaled in terms of flux in order to match the mean flux evolution fit $\bar{F}(z)$ from Oñorbe et al. (2017). This rescaling is a standard procedure for accounting for our lack of knowledge of the precise value of the metagalactic ionizing background photoionization rate.

In Figure A.2 we show a comparison of the mean flux values inferred from our data set (black squares), the values in Becker et al. (2013) (red) and the fit to diverse mean flux measurements from Oñorbe et al. (2017) (dashed line) which was used as a basis for rescaling the mean flux of simulated spectra in Chapter 2. Only pixels that were not flagged as metals, high column density absorbers, or bad pixels were used for the calculation of the mean flux in our data. When looking at the mean flux of the data, we observe that they scatter around the mean flux used in the simulations in the range $z = 2.0$-3.0.

To motivate the fact that we do not take into account uncertainties in the mean flux rescaling of our simulations at $2 < z < 3$, we ran our measurements at $z = 2.4$



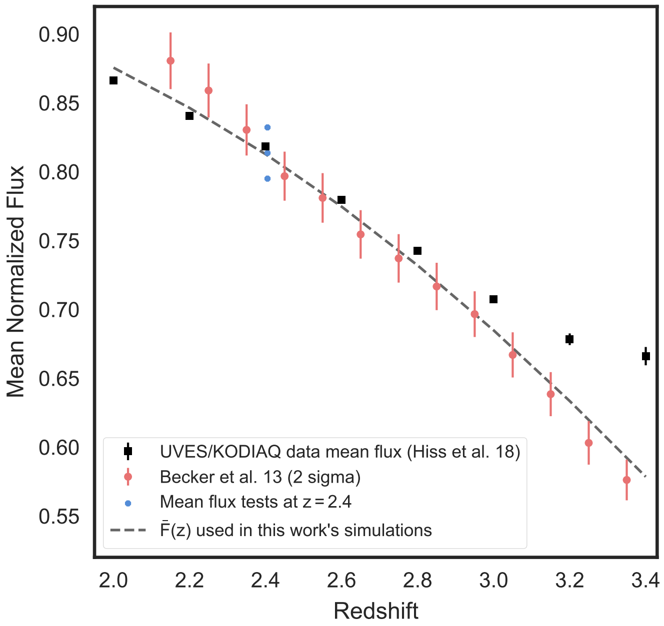


Figure A.2. Comparison of the mean flux evolution from Oñorbe et al. (2017) (dashed line, used as a basis for rescaling the mock skewers in Chapter 2), the measurements by Becker et al. (2013) (red points) and the mean flux of the data from § 2.1 at each redshift bin. The mean flux values used in for the test in Figure A.3 are shown in blue. Figure credit: Hiss et al. (2018).

for different values of the flux rescaling: $\bar{F}$, i.e. our measurement, and $\bar{F} \pm 2\sigma_{\bar{F}}$, where $\bar{F} = 0.8136$ is the value interpolated between the measurements of $\bar{F}$ by Becker et al. (2013) at $z = 2.35$ and 2.45. For the purpose of being conservative, the value of $\sigma_{\bar{F}}$ adopted is the error reported by Becker et al. (2013) at $z = 2.35$, $\sigma_{\bar{F}} = 0.0093$. These values are plotted as blue dots in Figure A.2. The corresponding $p(T_0, \gamma)$ measurements are shown in Figure A.3. Shifting $\bar{F}$ by 2σ results in a negligible shift of our final results at this redshift.

At our highest-redshift bins, $z = 3.2$ and 3.4 we observe a stronger discrepancy between the mean flux of our models and data. In order to directly examine the effect of this discrepancy on our measurements, we generated the models used in the calibration once again, with the difference that we rescaled the optical depths to match the mean flux values measured in the data at these redshifts. We then applied the calibration based on these new models to our cutoff fit results. The results are shown in Figure A.4. We observe that the calibrations at these redshifts are only slightly sensitive to this change, as our results basically do not change.

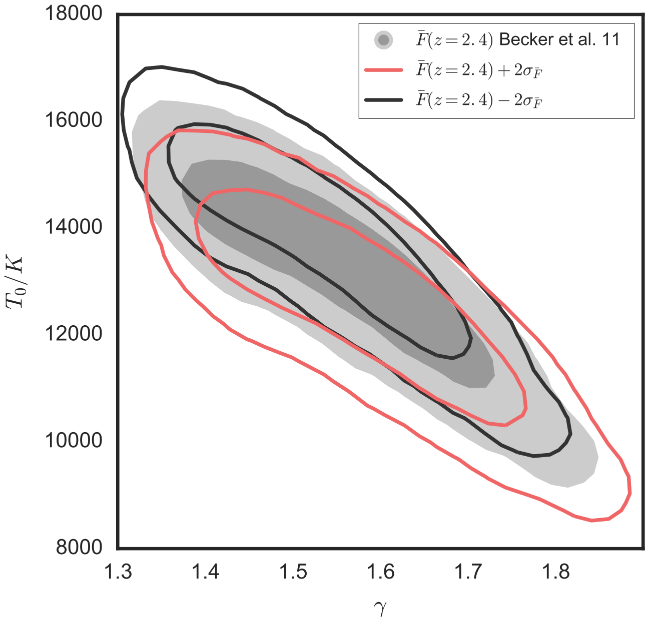
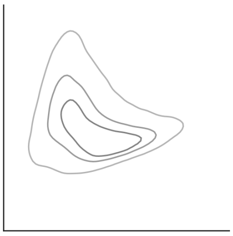


Figure A.3. Comparison of results at $z = 2.4$ for our data calibrated using simulations that are scaled to three different mean flux values: Becker et al. (2011) mean flux (gray filled contours), Becker et al. (2011) mean flux +2σ (red contour lines), and Becker et al. (2011) mean flux -2σ (black contour lines). The contours correspond to the 68% and 95% confidence regions. Figure credit: Hiss et al. (2018).



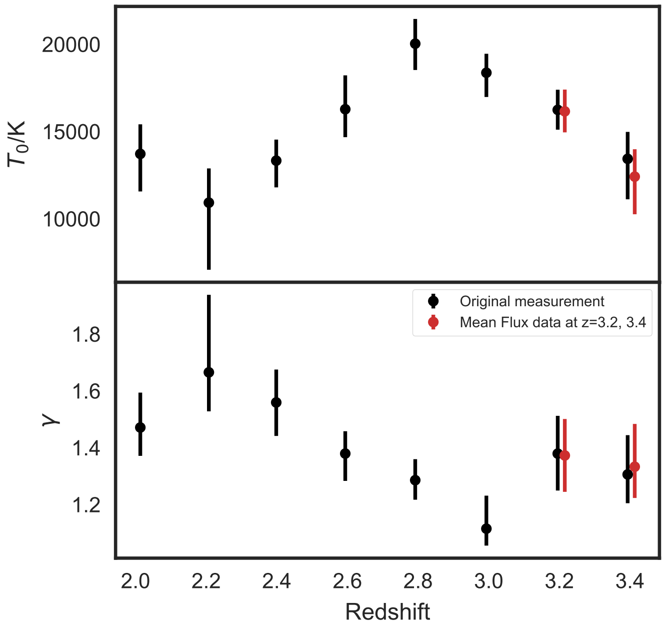


Figure A.4. Final marginalized T_0 and γ measurements after rescaling our models to match the mean flux of our data at $z = 3.2$ and 3.4 (red) compared to our original measurements (black). Figure credit: Hiss et al. (2018).

A.3. Calibration Bias

In § 2.3.3 we carry out a calibration that relates the cutoff parameters b_0 and Γ to the parameters of the TDR. Figure 2.15 shows a systematic dependency of the $\log T_0/\log b_0$ calibration on γ. The theoretically expected value of γ, as well as the value we measured, lie toward the upper end of the range used. Therefore calibrating with the mean value of the simulations might bias our results.

In order to quantify how significant this bias can be, we generated our $\log b_0$ vs $\log T_0$ calibration again, this time using models with γ around the original measurements (blue in Figure A.5). This is done to ensure that the mean γ of the simulations lie in a range comparable to our final measurements. The comparison between the measurements of T_0 with different calibrations is shown in Figure A.6. Using the whole grid biases the results slightly towards higher temperatures at low redshift. Generally speaking, the bias is smaller than 1σ. At $z = 3$ there is a larger discrepancy. Given that this is the redshift where we measure the smallest value of γ, the number of models that we could use for the test (See the 6th panel of Figure A.5) is relatively small and is likely the cause for this effect. Given the small effect we observe, we conclude that this bias is not strong in our measurements.

A.4. The $N_{\mathrm{H\,I}}$-Δ Relation in our Simulations

In § 2.3.2 we used the N_{HI}-Δ from Schaye (2001) in order to estimate $N_{\mathrm{HI},0}= N_{\mathrm{HI}}(\Delta = 1)$ for the cutoff fitting procedure. In this section we will demonstrate that this relation is reproduced in our simulations by comparing the Δ skewers in our simulations to the N_{HI} found by `VPFIT` in the corresponding flux skewers. This approach has one complication, namely that peculiar velocities in our simulations will shift lines in redshift space causing them to not match the density peaks in the simulations. To overcome this problem, we follow the approach by Schaye et al. (1999) and use the optical depth weighted density Δ_τ as a proxy for the actual overdensity field, as this will mitigate the fact that the absorption lines are shifted in redshift space and do not match the density field in real space.

In order to show how well our simulations reproduce the optical depth weighted density N_{HI}-Δ_τ relation, first we calculate $\Delta_\tau = \frac{\sum_i \tau_i \Delta_i}{\sum_i \tau_i}$ as defined in Bolton et al. (2014) for our simulation that uses the standard Haardt & Madau (2012) UV-background at $z = 2.4$. The sum was computed over all pixel positions i in real space that have a contribution to τ in redshift space.

We show that our simulation presents a clear relation by plotting Δ_τ versus the corresponding flux $F = \exp(-\tau)$ for all pixels. This relation is shown in Figure A.7 and illustrates that most of the gas is following a clear Δ_τ versus flux relation, indicating

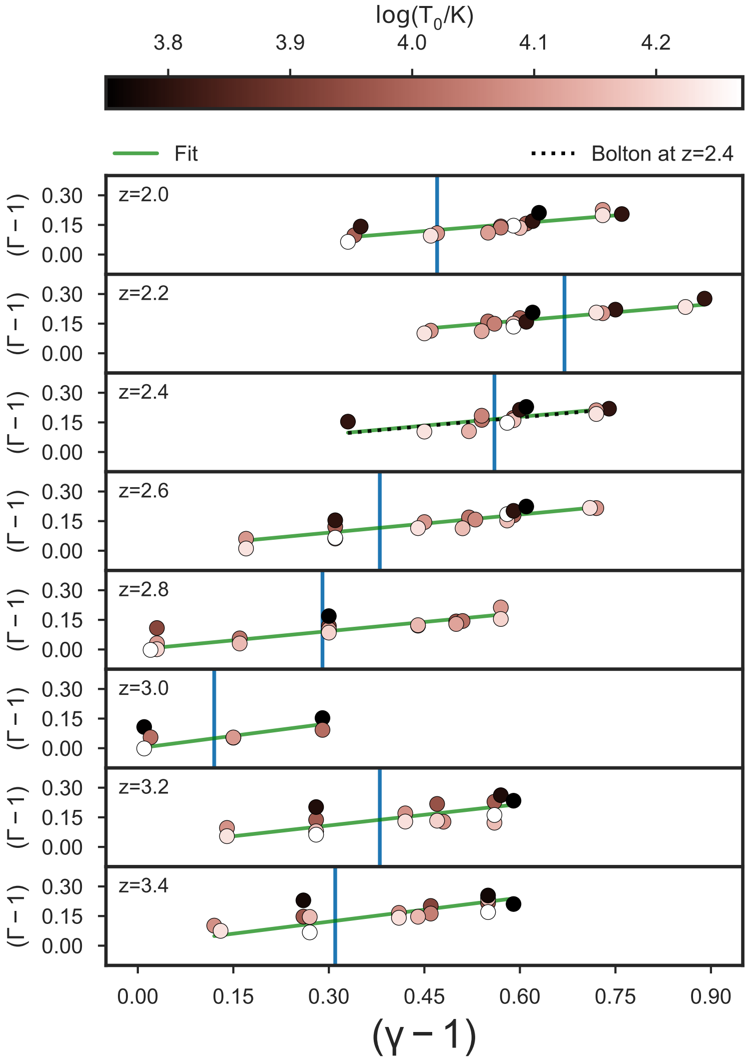


Figure A.5. Smaller grid chosen around final γ values used to test the systematic dependency of the $\log b_0$-$\log T_0$ calibration.

that Δ_τ is a good proxy for the overdensity field. To test if we reproduce eqn. 2.3 we match Δ_τ to $N_{\mathrm{H\,I}}$ found by `VPFIT` at the pixel position of the absorption line centers (See Figure A.8 for an example of how the matching was done). The matching of $N_{\mathrm{H\,I}}$ and Δ_τ is shown in Figure A.9.

We plot eqn. 2.3 using the parameters T_0, $\Gamma_{\mathrm{H\,I}}$, $z = 2.4$ from our simulation (black solid line). This figure illustrates that our simulations reproduce this relation very closely. A difference of 0.14 dex is seen in $N_{\mathrm{H\,I},0}$ when comparing the analytical relation to a linear fit to the median of the points in $\log \Delta_\tau$ bins. As will be shown in A.5, this

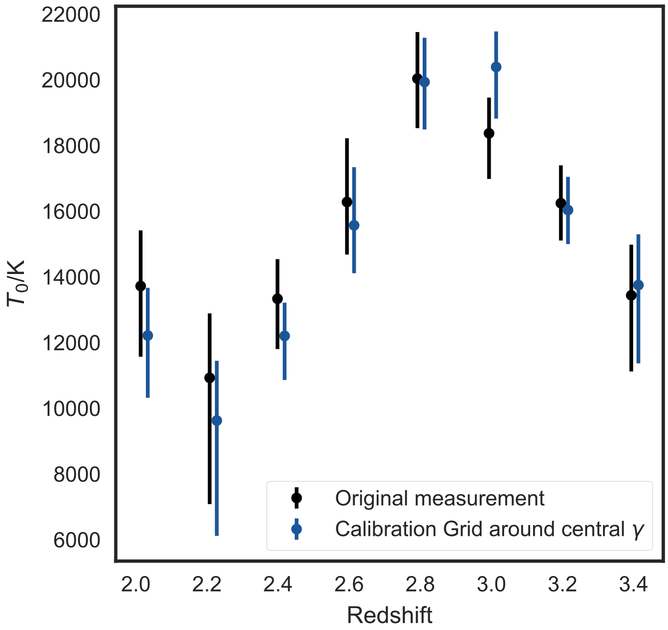


Figure A.6. Comparison of our final measurements of T_0 when using our complete model grid calibration and when using the models with γ shown in Figure A.5 in the $\log b_0$-$\log T_0$ calibration.

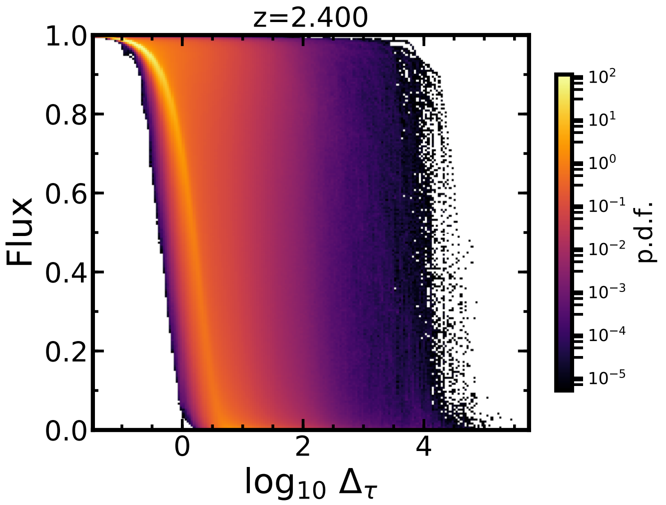


Figure A.7. The relation of Flux and Δ_τ in a Nyx simulation at $z = 2.4$. Figure credit: José Oñorbe).

difference does not play a role when the same $N_{\mathrm{HI},0}$ is applied both to the data and to the simulations when calibrating.

We also plot this relation for another simulation in our grid at $z = 2.4$, with slightly different thermal parameters (especially higher T_0 and slightly lower Γ_{HI} which are indicated in the figure) and Γ_{HI} (Figure A.10). While reproduce eqn. 2.3 reasonably well, the difference in $N_{\mathrm{HI},0}$ between the relation and a linear fit to the median of the points in $\log \Delta_\tau$ bins is of around 0.25 dex. The differences we would see in T_0 and γ due to a difference in $N_{\mathrm{HI},0}$ of 0.25 dex, are also expected to be small (compare to the effect of a difference of 0.2 dex in $N_{\mathrm{HI},0}$ on our measurements which we will discuss in Figure A.13).

To summarize, we quantified how the column densities of absorption profiles identified for two models in our simulations at $z = 2.4$ match the corresponding optical depth weighted density. We find that we reproduce eqn. 2.3 fairly well and that the difference between assuming the analytical form or calculating it directly from the diagram is of the order of 0.2 dex in $N_{\mathrm{HI},0}$ for the two examples. As will be discussed in the next section, this is a difference for which the effect on T_0 and γ was explicitly tested and showed to be not significant.

A.5. Effect of $N_{\mathrm{H\,I},0}$ on the Cutoff Measurements

The fact that we chose to use the analytic relation[1] from eqn. 2.3, was initially to motivate our choice of $N_{\mathrm{H\,I},0}$ in a somewhat physical way. Afterward we found that the choice does not matter when using our machinery, as will be shown in this section. In principle we could have chosen a fixed $N_{\mathrm{H\,I},0}$ value, say $\log N_{\mathrm{H\,I},0}$=13.2 throughout our whole redshift range, avoiding eqn. 2.3 altogether, and the results would remain unchanged in terms of T_0.

We argued that the choice of $N_{\mathrm{H\,I},0}$ does not affect our measurements significantly. This is the case because we choose the same values of $N_{\mathrm{H\,I},0}$ for computing b_0 in the data, as well as for the b_0-T_0 relation in the simulations. Using this relation as the basis of the calibration will neutralize the effect of $N_{\mathrm{H\,I},0}$ in the final result. We will elaborate on this in more detail in the text below. In the case of Rudie et al. (2012a) the choice of $N_{\mathrm{H\,I},0}$ made a critical difference, as the authors in this work did not use simulations as a basis for their calibration and instead chose a fixed analytical calibration.

The main motivation for choosing to normalize the column densities by $N_{\mathrm{H\,I},0}$ is that it makes b_0-T_0 calibration relation independent of γ, which simplifies the calibration. However, even if one chooses to normalize the column densities by a value which is not exactly $N_{\mathrm{H\,I},0}$, the impact on our analysis would be to simply add a scatter (and covariance with γ) to the calibration.

In order to investigate the effect of the choice of $N_{\mathrm{H\,I},0}$ on our measurements, first we calculated $N_{\mathrm{H\,I},0}$ with the analytical formula from Schaye (2001) (based on the value of $\Gamma_{\mathrm{H\,I}}$ and temperature of our hydrodynamic simulations). The results are shown in Figure 2.13 and in the rightmost panel of Figure 2.14. As discussed, our $N_{\mathrm{H\,I},0}(z = 2.4)$ from Nyx is about 0.3 dex higher than Bolton et al. (2014). As discussed in § 2.3.2, this discrepancy can be explained by the different values of $\Gamma_{\mathrm{H\,I}}$ used in their simulations and ours.

To understand the effects of $N_{\mathrm{H\,I},0}$ our inference, we compare the final T_0 and γ when adopting three different $N_{\mathrm{H\,I},0}$ curves. Note that in each case whatever $N_{\mathrm{H\,I},0}$ convention is chosen is used consistently in the data as well as in the simulations:

1. $N_{\mathrm{H\,I},0}$ based on hydrodynamic simulations (blue line in the right panel of Figure 2.14), shown in blue, i.e. our results from § 2.4.

2. $N_{\mathrm{H\,I},0}$ based on hydrodynamic simulations with an offset of +0.2 dex, shown in light-blue.

3. $N_{\mathrm{H\,I},0}$ based on hydrodynamic simulations with an offset of -0.2 dex, shown in dark-blue.

---

[1] Note that we explicitly show how well our simulations reproduce this relation in § A.4

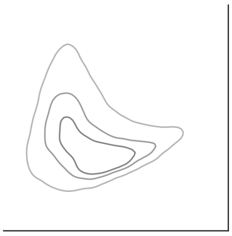
The effect of the change of $N_{\mathrm{H\,I},0}$ on the cutoff parameters is shown in Figure A.11. By definition, higher $N_{\mathrm{H\,I},0}$ results in higher b_0. As expected, the choice of $N_{\mathrm{H\,I},0}$, being simply a pivot point, should not affect the values of Γ and therefore not γ.

In Figure A.12 we illustrate how the change of ± 0.2 dex in $N_{\mathrm{H\,I},0}$ affects the cutoff fitting of our simulations, in terms of the calibration values D, C and κ. Here we see that the values D and C that parametrize the relation between $\log T_0$ and $\log b_0$ in the simulations also change with the choice of $N_{\mathrm{H\,I},0}$. This means, that we not only have $b_0(N_{\mathrm{H\,I},0})$, we also have $D(N_{\mathrm{H\,I},0})$ and $C(N_{\mathrm{H\,I},0})$, i.e. the relations (shown in Figure 2.15) will change with the choice of $N_{\mathrm{H\,I},0}$. Interestingly, we see that the variance in the calibration becomes smaller when increasing $N_{\mathrm{H\,I},0}$, but this does not affect our final results as our uncertainty in T_0 is dominated by the bootstrap estimated uncertainty in b_0.

The corresponding results (cutoff plus calibration) on the thermal parameters is shown in Figure A.13. We observe that the T_0 measurements are consistent with each other for all three conventions adopted for $N_{\mathrm{H\,I},0}$. The choice of $N_{\mathrm{H\,I},0}$ changes the calibration factors D and C along the degeneracy direction, compensating for the different choices of $N_{\mathrm{H\,I},0}$. The applied calibration results in Figure A.13 illustrating that the choice of $N_{\mathrm{H\,I},0}$ does not affect our final results if applied self-consistently to data and simulations (as long as the variance in the calibration is sub-dominant in the final error budget). In other words, the calibration compensates for the shift in $N_{\mathrm{H\,I},0}$.

An exact measurement of $N_{\mathrm{H\,I},0}$ would be necessary if we wanted to carry out an absolute calibration of b_0, i.e. if we were directly linking the model calibrations to something like $b_0 \propto \sqrt{T_0}$ which can only be done when b_0 actually corresponds to ρ_0. For a T_0 measurement, what we are doing currently is sufficient, as we are calibrating a correlation between b_0 and T_0 in simulations and applying this calibration to the data using the same $N_{\mathrm{H\,I},0}$, i.e. not necessarily expecting b_0 to correspond to the mean density.

Note that the different choices of $N_{\mathrm{H\,I},0}$ lead to more or less scatter in the $\log T_0$-$\log b_0$ calibration and that the choice of $N_{\mathrm{H\,I},0}$ becomes important only if this scatter is large. Therefore we argue that our measurements are in essence insensitive to the choice of $N_{\mathrm{H\,I},0}$.



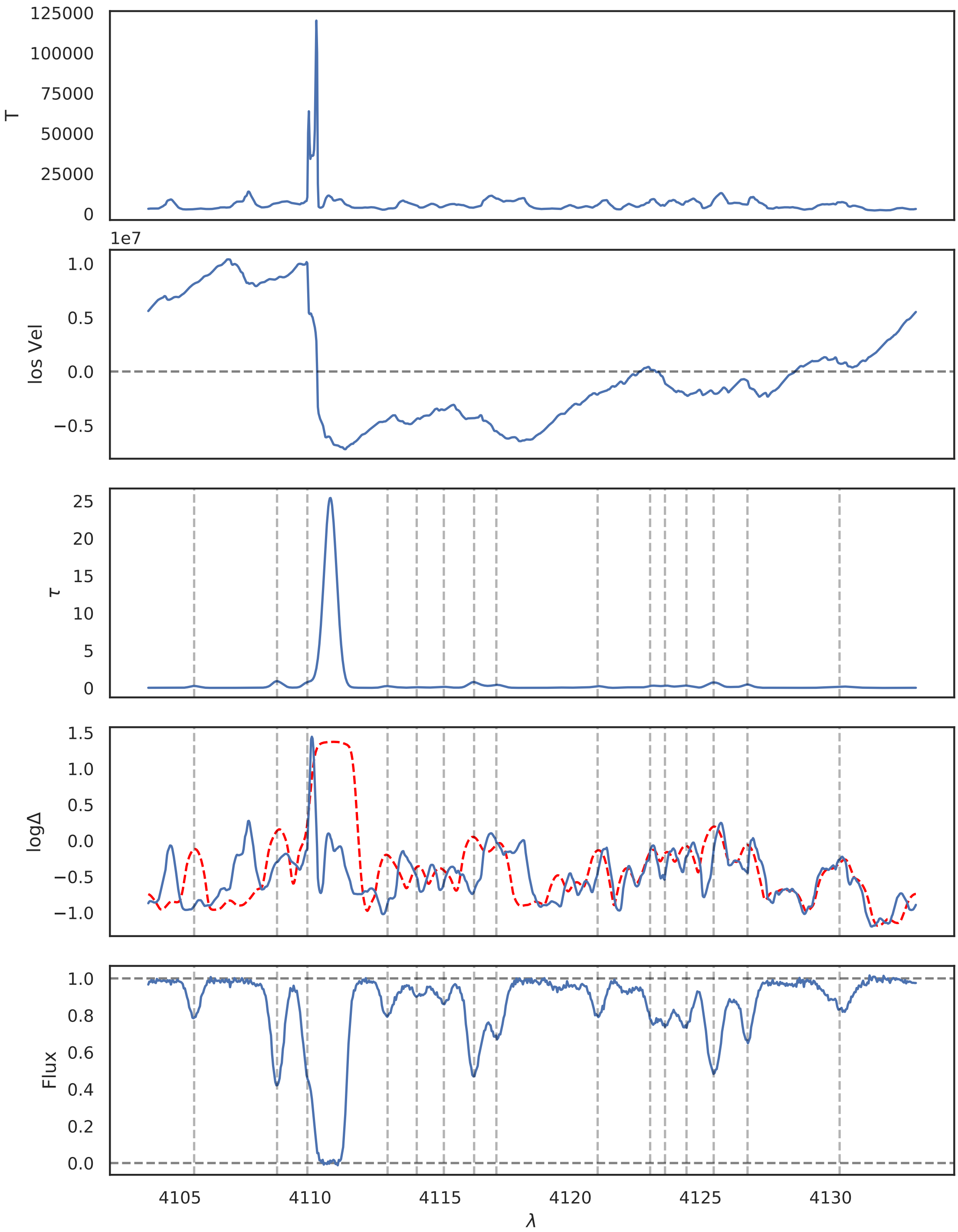
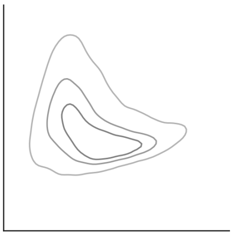


Figure A.8. Comparison of different physical properties of one skewer in our simulation. In descending order: temperature, line of sight velocity, optical depth, real space overdensity Δ (blue solid) compared to Δ_t (red dashed) and the Flux skewer that was fitted by VPFIT. The position of absorbers in redshift space with $12.5 < \log N_{\mathrm{HI}} < 14.5$ are shown as dashed lines.



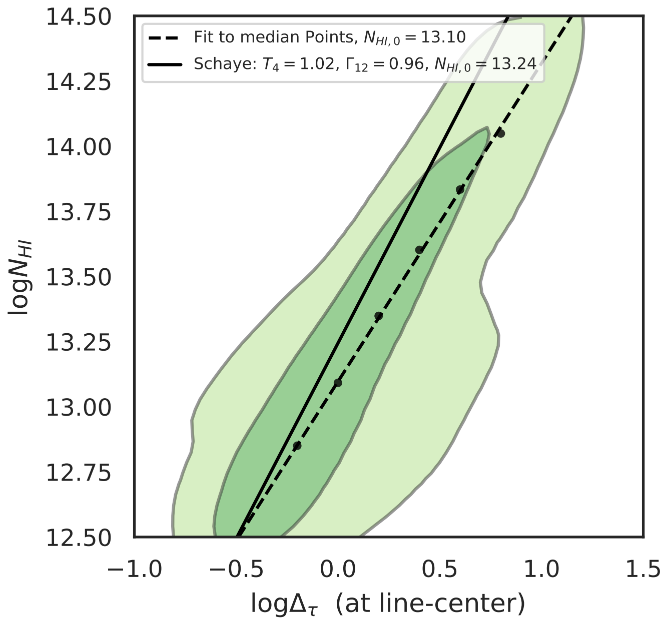


Figure A.9. Δ_τ versus $N_{\mathrm{H\,I}}$ relation for the simulation with thermal history using the Haardt & Madau (2012) UVB at $z = 2.4$.

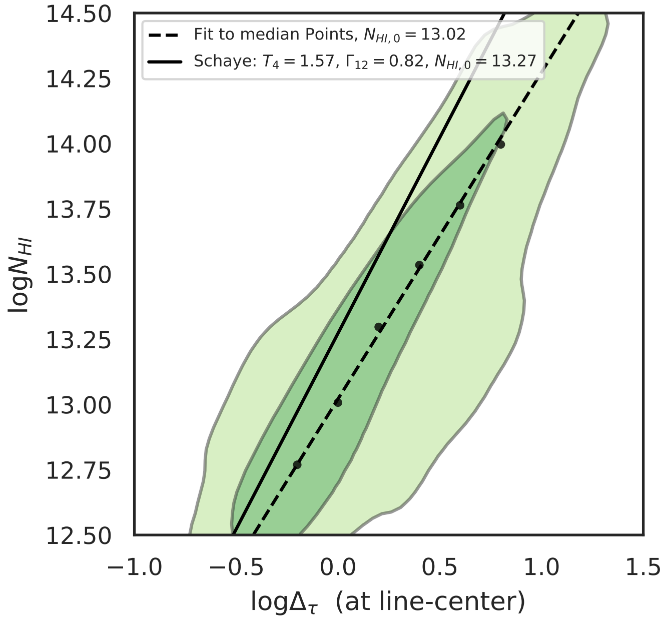


Figure A.10. Δ_τ versus $N_{\mathrm{H\,I}}$ relation for a simulation with a different thermal history at $z = 2.4$.

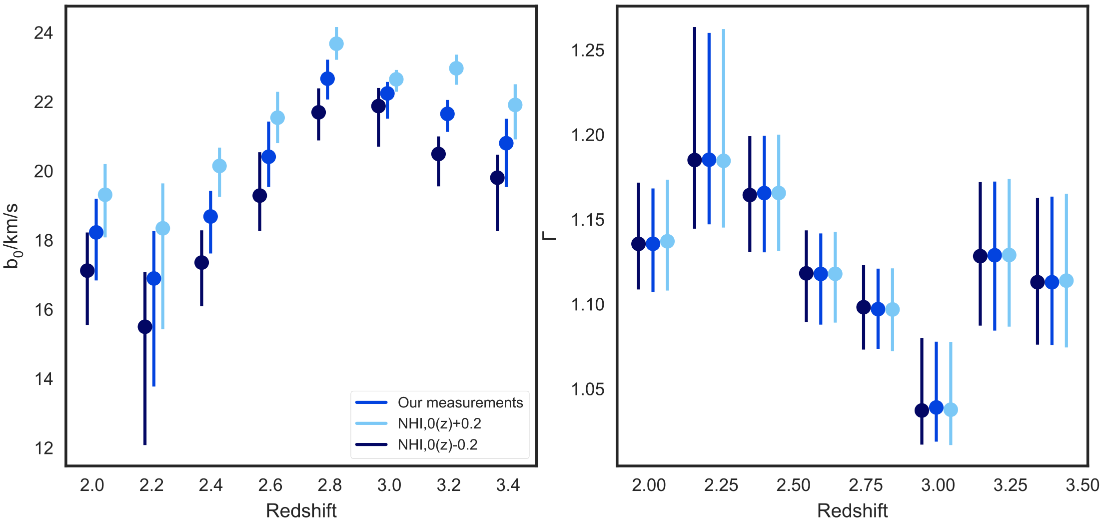


Figure A.11. Effect of the choice of $N_{\mathrm{H\,I},0}$ on the cutoff fit parameters. Points have been shifted slightly in redshift for visualization purposes.

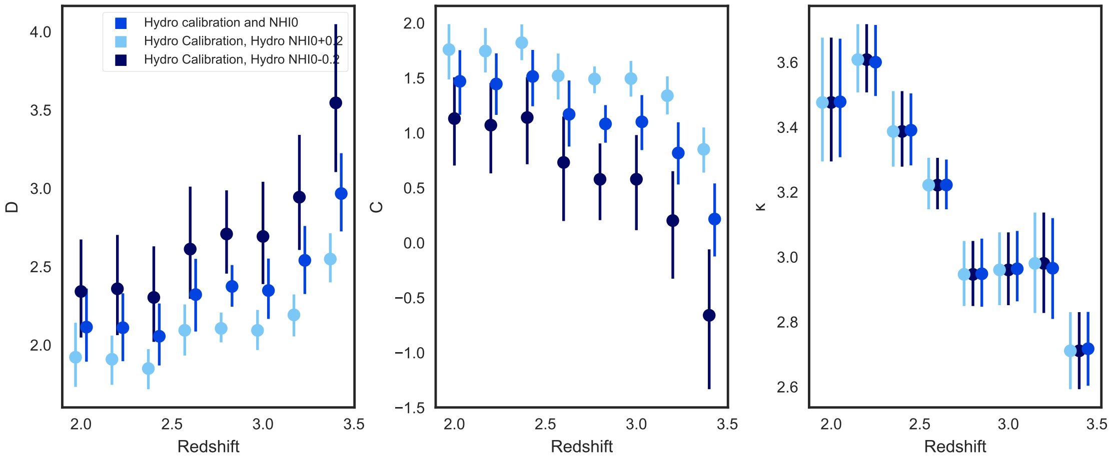


Figure A.12. Effect of the choice of $N_{\mathrm{H\,I},0}$ on the underlying calibration. Points have been shifted slightly in redshift for visualization purposes.

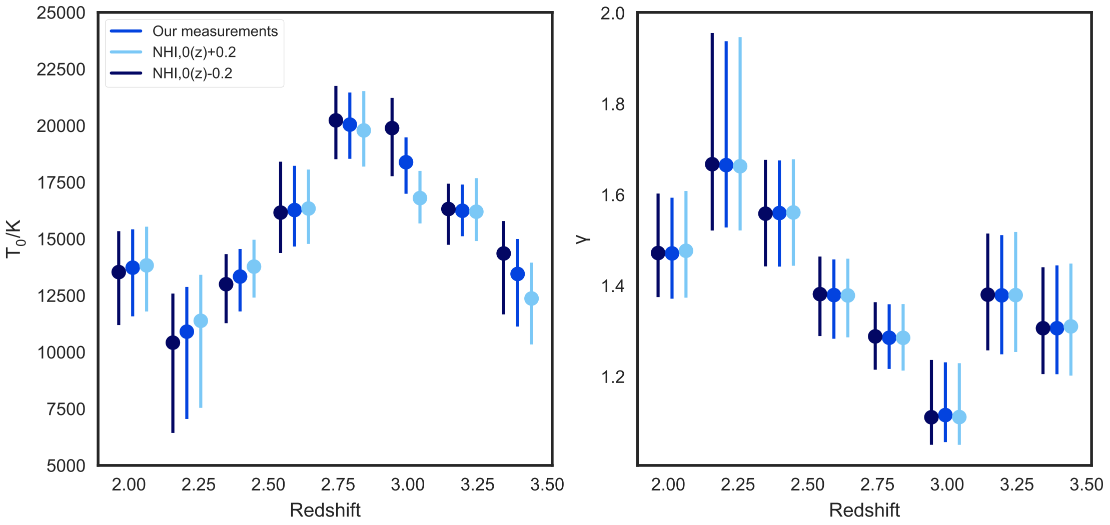


Figure A.13. Effect of the choice of $N_{\rm H\,I,0}$ on our final results. Points have been shifted slightly in redshift for visualization purposes.

B. Modeling the Full b-$N_{\mathrm{H\,I}}$ Distribution

B.1. Thermal Sensitivity Animations

In § 1.5.1 we illustrate in Figure 1.8 how changing the TDR parameters of DM-only models affects the shape of the b-$N_{\mathrm{H\,I}}$ distribution. This thesis comes with a special feature illustrating the sensitivity of the b-$N_{\mathrm{H\,I}}$ distribution to the parameters of the TDR: by flipping the pages of this thesis the reader can watch two animations showing b-$N_{\mathrm{H\,I}}$ distributions (now based on the hydrodynamic simulations from § 3.4.1) as a function of the parameters of the TDR. Flipping the pages and looking at the bottom right plots will show an animation of the effect of changing γ from 0.9 to 1.9 and back. Flipping the pages and looking at the bottom left plots will show an animation of the effect of changing T_0 from 4300 K to 18000 K and back.

B.2. Choice of Emulation Hyperparameters

B.2.1. Emulator Smoothing Length h_l

To motivate the choice of $h_l = 0.2$ for our emulator smoothing length for the DM-only models (see § 3.2.3), we compare the true PDF for a model with $\log T_0 = 4.128$ and $\gamma = 1.4165$ to the emulated PDF at the same thermal parameters using different smoothing lengths in Figure B.1. This particular model is not included in the emulator building process, but it is part of our test grid (see Figure 3.2), and was chosen to lie as far away from grid points as possible. We show the true PDF, i.e. the one computed directly from the b-$N_{\mathrm{H\,I}}$ distribution, in the top left panel and the difference between emulated and real PDF for different h_l in the other panels. Essentially, the emulated PDF differs substantially from the true one when choosing very small (smaller than grid separation, i.e. emulator does not correlate neighboring models) and very large h_l (factor of >3 grid separation). The emulator shows a stable performance in the intermediate range $0.03 \leq h_l \leq 0.3$, which implies that the choice of $h_l = 0.2$ is adequate. Note that this example shows the worst-case scenario where the emulated PDF is the farthest away from the grid points and the interpolation has the highest uncertainty.

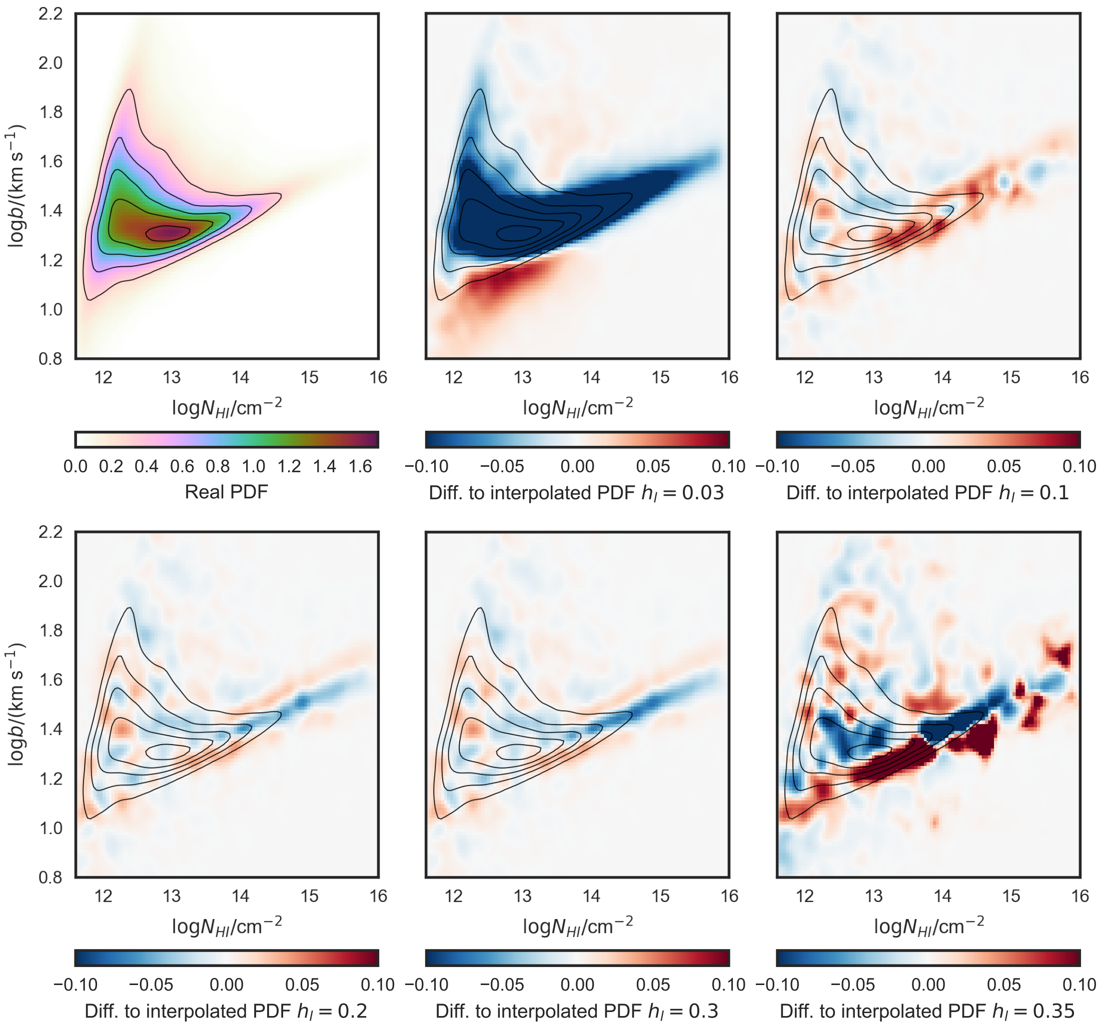
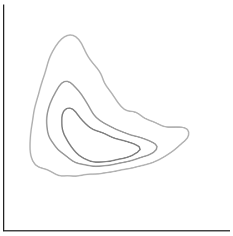


Figure B.1. Difference between true and emulated maps for a model with $\log T_0 = 4.128$ and $\gamma = 1.4165$ (not included in the DM-only emulation grid and maximally far away from points in the grid). The emulated PDFs were constructed from emulators using different smoothing lengths. Using a smoothing length that is too small results in an interpolation that does not take into account close grid points, while a large smoothing length introduces artifacts. We observe small fluctuations in comparison with the true PDF for intermediate $0.03 < h_l < 0.3$ for our DM-only emulation scheme. Figure credit: Hiss et al. (2019).



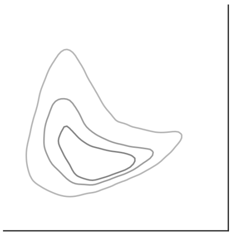
B.2.2. White-Noise Contribution σ_n

In § 3.4.1 we state that we chose the value $\sigma_n = 0.01$ for the hydrodynamic simulation grid based on visual inspection, because we observed clear interpolation artifacts in a few places inside the thermal grid when adopting no white-noise contribution. To explore the effect of this choice, we show in Figure B.2 one example with $\log T_0 = 3.9$ and $\gamma = 1.19$ that generated such artifacts. The top left panel shows the color-coded map and contours for the choice used in this study. All other panels represent different choices of white-noise contribution. Note that this choice of $\log T_0$ and γ represents the worst case of artifacts we encountered within the grid, and it corresponds to a location where the interpolation covers a substantial gap in parameter space. Unfortunately, we do not have the option of generating extra models as we did with the DM-only simulations, i.e. the true PDF at this grid position is unknown, but Figure B.2 indicates that (in this worst-case scenario) the general shape of the emulated b-$N_{\mathrm{H\,I}}$ distribution does not present artifacts for $\sigma_n > 0.005$ and keeps its general shape until σ_n is large (> 0.1) and that the interpolation has so much freedom in the grid points that the shape of the b-$N_{\mathrm{H\,I}}$ distribution loses information about the thermal state of the gas.

B.3. Effect of Different Data Subsampling Methods

As stated in § 3.4.2, we chose to draw 200 absorbers randomly from the dataset from § 2.1 at $z = 2$, because the models used to construct the b-$N_{\mathrm{H\,I}}$ distribution PDFs have a mixed SNR with a distribution based on our data. This approach could pose a problem, as random picking across the full dataset essentially removes correlations between absorbers in the same spectra. We showed in § 3.3.2, using DM-only simulations, that our inference is robust in the case of a fixed SNR and mock datasets composed of eight randomly drawn skewers, i.e. correlations are included and the SNR does not affect our inference test. To determine wether these effects play a role in the measurement presented in § 3.4.3, one should investigate the effects of picking random QSO sightlines instead of random absorbers, given that our likelihood is agnostic to correlations between absorption lines. In the following paragraphs we will explore both approaches.

To test wether our inference is influenced by randomly choosing absorbers, we generated another 200 realizations of 200 randomly chosen absorbers from the full dataset and carried out the same inference as in § 3.3.1. Note that while the same absorbers are present in different realizations, absorbers are picked without replacement such that the same absorber does not appear more than once in each individual realization. As a measure for how consistent the measurements of all these realizations are with each other, given that we do not know the true value, we compare the measurements of each



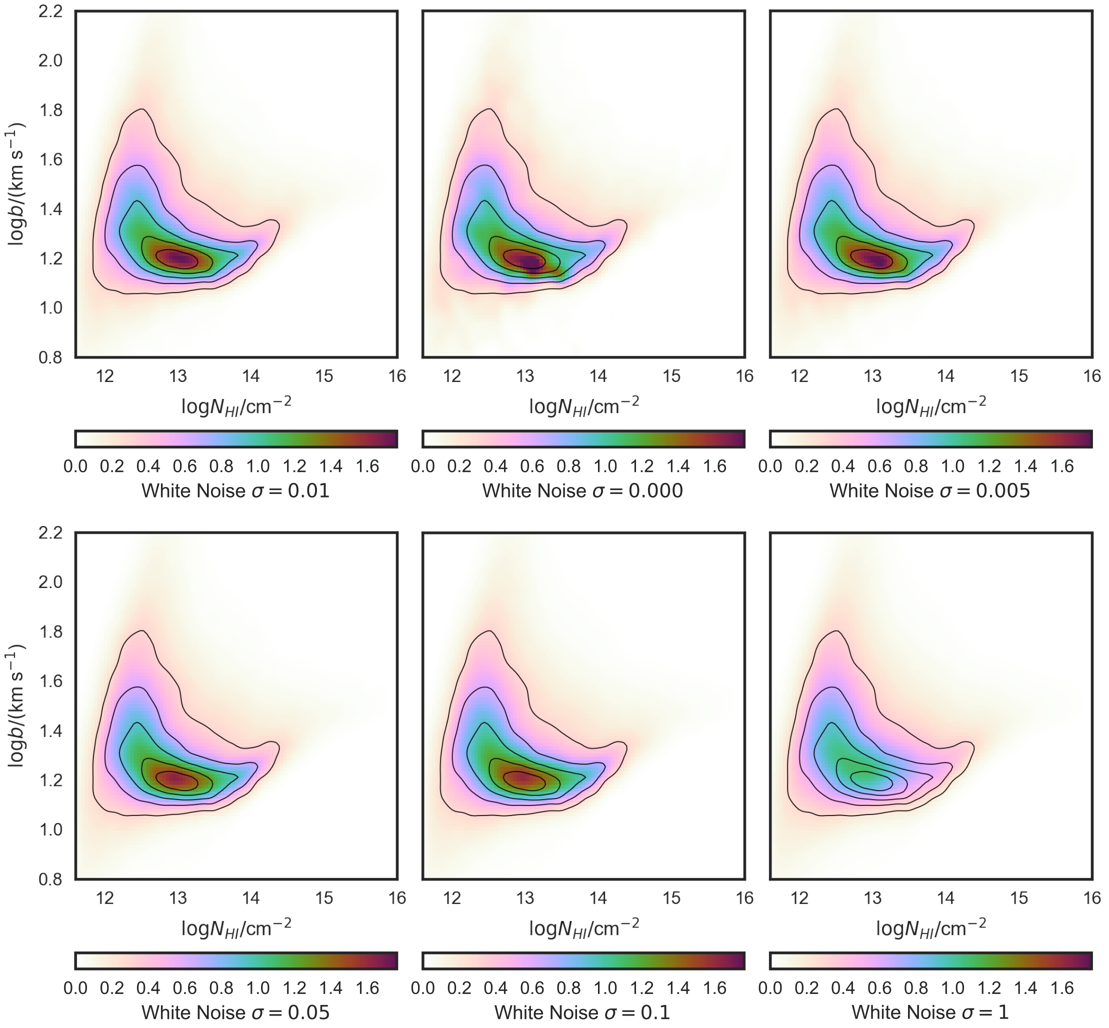
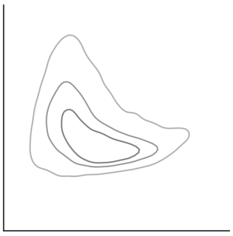


Figure B.2. Emulated b-$N_{\mathrm{H\,I}}$ distribution for $\log T_0 = 3.9$ and $\gamma = 1.19$ using the hydrodynamic grid for different values of the white-noise term σ_n. The emulated PDF resulting from an emulation using our fiducial choice of $\sigma_n = 0.01$ is shown in the top left panel as a color-coded map and corresponding contours. All other panels show the b-$N_{\mathrm{H\,I}}$ distribution, but emulated using different white-noise contributions. The contours of our fiducial choice are shown for comparison in all panels. This figure illustrates that allowing no freedom for the interpolation at the grid points results in interpolation artifacts in this particular position (between grid points). Additionally, allowing too much freedom results in loss of information about the thermal state. Figure credit: Hiss et al. (2019).



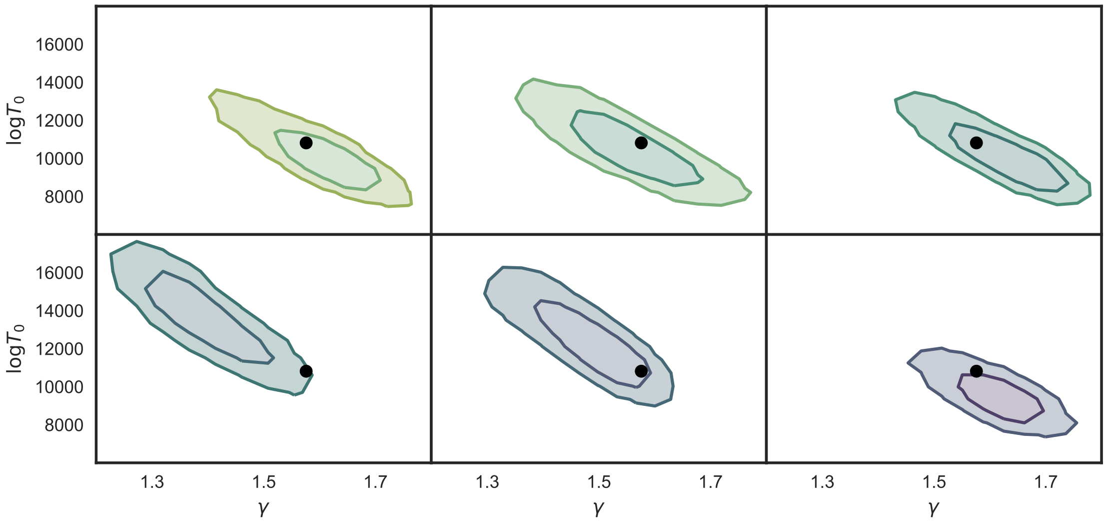


Figure B.3. The 1σ and 2σ contours for 6 measurement realizations (out of 200), each consisting of 200 unique randomly chosen absorbers from the dataset from § 2.1 at $z = 2$. The black circle illustrates the median of the measurements using the full dataset. Figure credit: Hiss et al. (2019).

realization to the measurement using the full dataset presented in § 3.4.3. We observe that the measurements from the full dataset ($\log T_0 = 4.034$ and $\gamma = 1.576$) are within the 1σ contour of the 2D posteriors of these realizations 65% (129/200) of the time and within the 2σ contour of the 2D posterior 96% (192/200) of the time. This implies that our inference is consistent in the limit of random realizations based on absorbers. For illustration, the posteriors for six realizations are shown in Figure B.3. For reference we also plot the measurement from the full dataset as a black circle.

We ran a similar test, this time choosing random QSO sightlines instead of random absorbers. Due to metal line masking, at $z = 2$ each QSO in our sample contributes with ~ 100 absorbers, which means that we would carry out a measurement using around two sightlines each time (see discussion in § 3.3.1). To test wether we achieve results that are consistent with the full dataset, we carried out this experiment using 11 quasars that span or nearly span the pathlength within $1.9 < z < 2.1$, which results in 55 unique pairs of quasars and therefore measurement realizations. We observe that the reference values measured using the full dataset are within the 1σ contours of the 2D MCMC posteriors of these realizations about 33% (18/55) of the time. This implies that there is some bias associated with choosing QSO sightlines randomly instead of absorbers. Note that we do not have sufficient statistics to quantify the behavior of the 95% contours with a sample size of 55 realizations.

One possible reason for failing this inference test when choosing the pairs of QSO

sightlines is the fact that we are choosing nonrepresentative SNR values by picking random QSOs and comparing their b-$N_{\rm H\,I}$ distributions to models that were constructed to match the SNR distribution of the whole dataset. The proper approach to remove a possible SNR bias would be to generate a set of models with the matching SNR for each data subsample separately, i.e. generate a set of forward models for every quasar pair in the example above. This approach would require applying VPFIT to our full model grid and recreating a b-$N_{\rm H\,I}$ distribution emulator for every MCMC posterior we wish to generate. We have considered this approach, but we concluded that it implies a significant computational effort, given that the current calculations are already extremely resource consuming when done once. Additionally, real physical sightline-to-sightline variations in the TDR could also contribute to the poor performance of this inference test. If present, these variations would mean that subsampling by choosing random absorbers essentially results in a measurement of the average TDR in that specific subsample.

List of the Author's Publications

Here I present the full list of publications I was involved in prior to and during my PhD. This list includes papers, coauthored papers, database entries and proceedings. Chapter 2 of this thesis is based on Hiss et al. (2018) and Chapter 3 is based on Hiss et al. (2019).

Hiss et al. (2019)
First author, refereed paper

Hiss et al. (2018)
First author, refereed paper

Hiss et al. (2018)
Online database with DOI

Walther et al. (2018)
Coauthor, refereed paper

Csépány et al. (2017)
Coauthor, refereed paper

Kohler & Hiss (2015)
Conference proceedings

Please refer to the bibliography for more details.

Bibliography

Almgren, A. S., Bell, J. B., Lijewski, M. J., Lukić, Z., & Van Andel, E. 2013, ApJ, 765, 39

Alpher, R. A., & Herman, R. 1948, Nature, 162, 774

Alsing, J., Charnock, T., Feeney, S., & Wand elt, B. 2019, MNRAS, 1888

Alsing, J., Wandelt, B., & Feeney, S. 2018, MNRAS, 477, 2874

Ambikasaran, S., Foreman-Mackey, D., Greengard, L., Hogg, D. W., & O'Neil, M. 2016, TPAMI, 38, 252

Bañados, E., Venemans, B. P., Mazzucchelli, C., et al. 2018, Nature, 553, 473

Becker, G. D., & Bolton, J. S. 2013, MNRAS, 436, 1023

Becker, G. D., Bolton, J. S., Haehnelt, M. G., & Sargent, W. L. W. 2011, MNRAS, 410, 1096

Becker, G. D., Bolton, J. S., Madau, P., et al. 2015, MNRAS, 447, 3402

Becker, G. D., Hewett, P. C., Worseck, G., & Prochaska, J. X. 2013, MNRAS, 430, 2067

Boera, E., Becker, G. D., Bolton, J. S., & Nasir, F. 2018, ArXiv e-prints, arXiv:1809.06980

Boera, E., Murphy, M. T., Becker, G. D., & Bolton, J. S. 2014, MNRAS, 441, 1916

Bolton, J. S., Becker, G. D., Haehnelt, M. G., & Viel, M. 2014, MNRAS, 438, 2499

Bolton, J. S., Viel, M., Kim, T.-S., Haehnelt, M. G., & Carswell, R. F. 2008, MNRAS, 386, 1131

Bosman, S. E. I., Fan, X., Jiang, L., et al. 2018, MNRAS, 479, 1055

Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2015, ApJ, 803, 34

Bryan, G. L., & Machacek, M. E. 2000, ApJ, 534, 57

Carswell, R. F., & Webb, J. K. 2014, VPFIT: Voigt profile fitting program, Astrophysics Source Code Library, ascl:1408.015

Carswell, R. F., Webb, J. K., Cooke, A. J., & Irwin, M. J. 2014, RDGEN: Routines for data handling, display, and adjusting, Astrophysics Source Code Library, ascl:1408.017

Compostella, M., Cantalupo, S., & Porciani, C. 2013, MNRAS, 435, 3169

—. 2014, MNRAS, 445, 4186

Croft, R. A. C., Weinberg, D. H., Katz, N., & Hernquist, L. 1998, ApJ, 495, 44

Csépány, G., van den Ancker, M., Ábrahám, P., et al. 2017, A&A, 603, A74

Dall'Aglio, A., Wisotzki, L., & Worseck, G. 2008, AAP, 491, 465

D'Aloisio, A., McQuinn, M., & Trac, H. 2015, ApJL, 813, L38

Danforth, C., Pieri, M., Shull, J. M., et al. 2013, in American Astronomical Society Meeting Abstracts, Vol. 221, American Astronomical Society Meeting Abstracts #221, 245.04

Danforth, C. W., Keeney, B. A., Tilton, E. M., et al. 2016, ApJ, 817, 111

Davies, F. B., Furlanetto, S. R., & Dixon, K. L. 2017, MNRAS, 465, 2886

Davies, F. B., Hennawi, J. F., Bañados, E., et al. 2018, ApJ, 864, 142

de Graaff, A., Cai, Y.-C., Heymans, C., & Peacock, J. A. 2019, A&A, 624, A48

Dekker, H., D'Odorico, S., Kaufer, A., Delabre, B., & Kotzlowski, H. 2000, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 4008, Optical and IR Telescope Instrumentation and Detectors, ed. M. Iye & A. F. Moorwood, 534

Dixon, K. L., & Furlanetto, S. R. 2009, ApJ, 706, 970

Dixon, K. L., Furlanetto, S. R., & Mesinger, A. 2014, MNRAS, 440, 987

Draine, B. T. 2011, Physics of the interstellar and intergalactic medium (Princeton University Press), princeton series in astrophysics

Eilers, A.-C., Davies, F. B., & Hennawi, J. F. 2018, ApJ, 864, 53

Fan, X., Strauss, M. A., Becker, R. H., et al. 2006, AJ, 132, 117

Faucher-Giguère, C.-A., Lidz, A., Hernquist, L., & Zaldarriaga, M. 2008a, ApJ, 688, 85

Faucher-Giguère, C.-A., Prochaska, J. X., Lidz, A., Hernquist, L., & Zaldarriaga, M. 2008b, ApJ, 681, 831

Finkelstein, S. L., Ryan, Russell E., J., Papovich, C., et al. 2015, ApJ, 810, 71

Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306

Fumagalli, M., O'Meara, J. M., Prochaska, J. X., Kanekar, N., & Wolfe, A. M. 2014, MNRAS, 444, 1282

Furlanetto, S. R., & Oh, S. P. 2008, ApJ, 681, 1

Gaikwad, P., Srianand, R., Choudhury, T. R., & Khaire, V. 2017, MNRAS, 467, 3172

Garzilli, A., Bolton, J. S., Kim, T.-S., Leach, S., & Viel, M. 2012, MNRAS, 424, 1723

Garzilli, A., Theuns, T., & Schaye, J. 2015, MNRAS, 450, 1465

—. 2018, ArXiv e-prints, arXiv:1808.06646

Gnedin, N. Y., & Hui, L. 1996, ApJL, 472, L73

—. 1998, MNRAS, 296, 44

Goodman, J., & Weare, J. 2010, CAMCoS, 5, 65

Gunn, J. E., & Peterson, B. A. 1965, ApJ, 142, 1633

Haardt, F., & Madau, P. 2012, ApJ, 746, 125

Habib, S., Heitmann, K., Higdon, D., Nakhleh, C., & Williams, B. 2007, Phys. Rev. D, 76, arXiv:astro-ph/0702348

Haehnelt, M. G., & Steinmetz, M. 1998, MNRAS, 298, L21

Heitmann, K., Higdon, D., Nakhleh, C., & Habib, S. 2006, ApJL, 646, L1

Herbert-Fort, S., Prochaska, J. X., Dessauges-Zavadsky, M., et al. 2006, PASP, 118, 1077

Hinshaw, G., Larson, D., Komatsu, E., et al. 2013, ApJS, 208, 19

Hiss, H., Walther, M., Hennawi, J., et al. 2018, Results from Voigt profile fitting from Hiss et al. 2018

Hiss, H., Walther, M., Hennawi, J. F., et al. 2018, ApJ, 865, 42

Hiss, H., Walther, M., Oñorbe, J., & Hennawi, J. F. 2019, ApJ, 876, 71

Hui, L., & Gnedin, N. Y. 1997, MNRAS, 292, 27

Hui, L., & Rutledge, R. E. 1999, ApJ, 517, 541

Katz, N., Hernquist, L., & Weinberg, D. H. 1992, ApJL, 399, L109

Khaire, V., & Srianand, R. 2019, MNRAS, 484, 4174

Khaire, V., Srianand, R., Choudhury, T. R., & Gaikwad, P. 2016, MNRAS, 457, 4051

Khaire, V., Walther, M., Hennawi, J. F., et al. 2019, MNRAS, 486, 769

Kim, T.-S., Bolton, J. S., Viel, M., Haehnelt, M. G., & Carswell, R. F. 2007, MNRAS, 382, 1657

Kim, T.-S., Viel, M., Haehnelt, M. G., Carswell, R. F., & Cristiani, S. 2004, MNRAS, 347, 355

Kirkman, D., Tytler, D., Suzuki, N., et al. 2005, MNRAS, 360, 1373

Kohler, R., & Hiss, H. 2015, in Cambridge Workshop on Cool Stars, Stellar Systems, and the Sun, Vol. 18, 18th Cambridge Workshop on Cool Stars, Stellar Systems, and the Sun, ed. G. T. van Belle & H. C. Harris, 237

Kollmeier, J. A., Weinberg, D. H., Oppenheimer, B. D., et al. 2014, ApJL, 789, L32

Kulkarni, G., Hennawi, J. F., Oñorbe, J., Rorai, A., & Springel, V. 2015, Apj, 812, 30

Lamberts, A., Chang, P., Pfrommer, C., et al. 2015, ArXiv e-prints, arXiv:1502.07980

Lee, K.-G., Hennawi, J. F., Spergel, D. N., et al. 2015, ApJ, 799, 196

Lehner, N., O'Meara, J. M., Fox, A. J., et al. 2014, ApJ, 788, 119

Lidz, A., Faucher-Giguère, C.-A., Dall'Aglio, A., et al. 2010, ApJ, 718, 199

Loeb, A., & Barkana, R. 2001, ARAA, 39, 19

Lukić, Z., Stark, C. W., Nugent, P., et al. 2015, MNRAS, 446, 3697

Lynds, R. 1971, Apj, 164, L73

Madau, P., & Haardt, F. 2015, ApJL, 813, L8

Madau, P., & Meiksin, A. 1994, ApJL, 433, L53

McDonald, P., Miralda-Escudé, J., Rauch, M., et al. 2001, ApJ, 562, 52

McDonald, P., Seljak, U., Burles, S., et al. 2006, ApJS, 163, 80

McGreer, I. D., Mesinger, A., & D'Odorico, V. 2015, MNRAS, 447, 499

McQuinn, M., Lidz, A., Zaldarriaga, M., et al. 2009, Apj, 694, 842

McQuinn, M., & Upton Sanderbeck, P. R. 2016, MNRAS, 456, 47

Meiksin, A., & White, M. 2001, MNRAS, 324, 141

Meiksin, A. A. 2009, Reviews of Modern Physics, 81, 1405

Mesinger, A. 2016, Understanding the Epoch of Cosmic Reionization: Challenges and Progress, Vol. 423

Miralda-Escudé, J., Haehnelt, M., & Rees, M. J. 2000, ApJ, 530, 1

Oñorbe, J., Hennawi, J. F., & Lukić, Z. 2017, ApJ, 837, 106

O'Meara, J. M., Lehner, N., Howk, J. C., et al. 2017, ArXiv e-prints, arXiv:1707.07905

—. 2015, ArXiv e-prints, arXiv:1505.03529

—. 2016, VizieR Online Data Catalog, 515

Palanque-Delabrouille, N., Yèche, C., Borde, A., et al. 2013, AAP, 559, A85

Papamakarios, G., Pavlakou, T., & Murray, I. 2017, arXiv e-prints, arXiv:1705.07057

Pfrommer, C., Broderick, A. E., Chang, P., Puchwein, E., & Springel, V. 2013, ArXiv e-prints, arXiv:1308.6284 [astro-ph.HE]

Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2014, AAP, 571, A16

Pritchard, J. R., & Loeb, A. 2012, Reports on Progress in Physics, 75, 086901

Puchwein, E., Bolton, J. S., Haehnelt, M. G., et al. 2015, MNRAS, 450, 4081

Puchwein, E., Haardt, F., Haehnelt, M. G., & Madau, P. 2019, MNRAS, 485, 47

Puchwein, E., Pfrommer, C., Springel, V., Broderick, A. E., & Chang, P. 2012, MNRAS, 423, 149

Ricotti, M., Gnedin, N. Y., & Shull, J. M. 2000, ApJ, 534, 41

Robertson, B. E., Ellis, R. S., Furlanetto, S. R., & Dunlop, J. S. 2015, ApJL, 802, L19

Rorai, A., Carswell, R. F., Haehnelt, M. G., et al. 2018, MNRAS, 474, 2871

Rorai, A., Hennawi, J. F., & White, M. 2013, ApJ, 775, 81

Rorai, A., Becker, G. D., Haehnelt, M. G., et al. 2017a, MNRAS, 466, 2690

Rorai, A., Hennawi, J. F., Oñorbe, J., et al. 2017b, Science, 356, 418

Rudie, G. C., Steidel, C. C., & Pettini, M. 2012a, ApJ, 757, L30

Rudie, G. C., Steidel, C. C., Trainor, R. F., et al. 2012b, ApJ, 750, 67

Schaye, J. 2001, Apj, 559, 507

Schaye, J., Theuns, T., Leonard, A., & Efstathiou, G. 1999, MNRAS, 310, 57

Schaye, J., Theuns, T., Rauch, M., Efstathiou, G., & Sargent, W. L. W. 2000, MNRAS, 318, 817

Seabold, S., & Perktold, J. 2010, in 9th Python in Science Conference

Silverman, B. W. 1986, Density estimation for statistics and data analysis

Sironi, L., & Giannios, D. 2014, ApJ, 787, 49

Smoot, G. F., Bennett, C. L., Kogut, A., et al. 1992, ApJL, 396, L1

Sorini, D., Oñorbe, J., Hennawi, J. F., & Lukić, Z. 2018, ApJ, 859, 125

Sorini, D., Oñorbe, J., Lukić, Z., & Hennawi, J. F. 2016, ApJ, 827, 97

Syphers, D., & Shull, J. M. 2014, ApJ, 784, 42

Tanimura, H., Aghanim, N., Douspis, M., Beelen, A., & Bonjean, V. 2019, A&A, 625, A67

Telikova, K., Balashev, S., & Shternin, P. 2018, ArXiv e-prints, arXiv:1806.01319

Theuns, T., Schaye, J., & Haehnelt, M. G. 2000, MNRAS, 315, 600

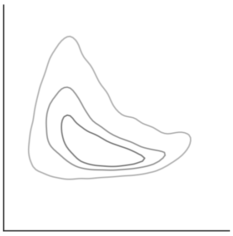
Theuns, T., Schaye, J., Zaroubi, S., et al. 2002, ApJl, 567, L103

Upton Sanderbeck, P. R., D'Aloisio, A., & McQuinn, M. J. 2016, MNRAS, 460, 1885

Viel, M., Bolton, J. S., & Haehnelt, M. G. 2009, MNRAS, 399, L39

Viel, M., Haehnelt, M. G., Bolton, J. S., et al. 2017, MNRAS, 467, L86

Viel, M., Haehnelt, M. G., & Springel, V. 2006, MNRAS, 367, 1655



Vogt, S. S., Allen, S. L., Bigelow, B. C., et al. 1994, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 2198, Instrumentation in Astronomy VIII, ed. D. L. Crawford & E. R. Craine, 362

Walther, M., Hennawi, J. F., Hiss, H., et al. 2018, ApJ, 852, 22

Walther, M., Oñorbe, J., Hennawi, J. F., & Lukić, Z. 2019, ApJ, 872, 13

Weinberg, D. H., Davé, R., Katz, N., & Kollmeier, J. A. 2003, in American Institute of Physics Conference Series, Vol. 666, The Emergence of Cosmic Structure, ed. S. H. Holt & C. S. Reynolds, 157

Weinberg, D. H., Hernsquit, L., Katz, N., Croft, R., & Miralda-Escudé, J. 1997, in Structure and Evolution of the Intergalactic Medium from QSO Absorption Line System, ed. P. Petitjean & S. Charlot, 133

White, M., Hernquist, L., & Springel, V. 2002, ApJ, 579, 16

Worseck, G., Davies, F. B., Hennawi, J. F., & Prochaska, J. X. 2018, ArXiv e-prints, arXiv:1808.05247

Worseck, G., Prochaska, J. X., McQuinn, M., et al. 2011, ApJl, 733, L24

Zaldarriaga, M., Hui, L., & Tegmark, M. 2001, ApJ, 557, 519

List of Acronyms

AGN active galactic nuclei

CMB Cosmic Microwave Background

COS Cosmic Origins Spectrograph

DELFI density-estimation likelihood-free inference

DM dark matter

DLA damped Lyα

GP Gaussian process

HIRES High Resolution Echelle Spectrometer

HST Hubble Space Telescope

IGM intergalactic medium

KDE Kernel Density Estimation

KODIAQ Keck Observatory Database of Ionized Absorbers toward QSOs

LD least absolute deviation

LLS Lyman limit systems

LS least squares

LSF line spread function

MCMC Markov chain Monte Carlo

MW Milky Way

NDE neural density estimation

PCA principal component analysis

PDF probability density function

PKP PCA decomposition of KDE estimates of a PDF

QSO quasi-stellar objects

SNR signal-to-noise ratio

STIS space telescope imaging spectrograph

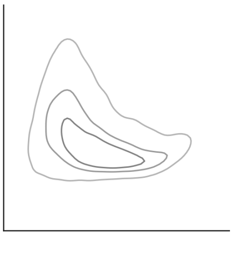
TDR temperature-density relation

THERMAL Thermal History and Evolution in Reionization Models of Absorption Lines

UV ultraviolet

UVB ultraviolet background

UVES Ultraviolet and Visual Echelle Spectrograph



Acknowledgments

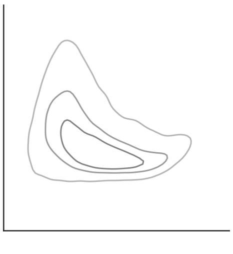
This thesis is dedicated to the Fäustle family, and especially to my friend and brother Markus Fäustle, who are the people that proved to me that love is out there, and that it is still alive and well. Thanks to my own family for making all of these pages a reality! Thanks to my mother Benícia M. Pará and my father Helmut W. Hiß for believing in my crazy dreams. I thank Antonia Fröhlich from the bottom of my heart for secretly buying me my first electric guitar, you saved my soul a million times. You are missed and I wouldn't be here without that help.

Thanks to all the members of the former office 217 at the Max Planck Institute for Astronomy: Sasa Tsatsi (*the* Morena), Paolo Bianchini, Hans Baehr (thanks for proofreading!), Sven Buder and Tobias Buck. May you all be blessed by Papa Francesco.

Special thanks to Stephen Christopher, Simon Walker, Andreas Finger and Andreas Pilder for being a boogle of incredible musicians I have the honor of sharing the stage with. Thanks to my dearest friends Armin Schwierk, Can Akin, Daniel Rosenblatt, Kerstin "Keki" Bojeczan, Sarah Wick, Ioannis Vasiliadis, Lyth Mashni, Nicolas Brod Bach. I regret if this thesis distanced me from you a little bit.

Special thanks to José Onõrbe for being a faithful advisor in the "dark ages" of my own little PhD universe and Joseph F. Hennawi, my supervisor, for keeping me motivated and pushing me to always go one step further and achieve the unthinkable. Thanks to Michael Walther, not only for all these years of day-to-day science support and exchange of Python wisdom, but also for the good time. Thanks to all members and former members of the ENIGMA group I had the honor to work with.

This thesis would not have been possible without the International Max Planck Research School for Astronomy and Cosmic Physics at the University of Heidelberg (IMPRS-HD) and especially without Christian Fendt's advice and support. Special thanks to Martin White for providing the collisionless DM simulations used in Chapters 2 and 3. Calculations presented in this thesis used the `hydra` and `draco` clusters of the Max Planck Computing and Data Facility (MPCDF, formerly RZG). MPCDF is a competence center of the Max Planck Society located in Garching (Germany). Some data presented in this work were obtained from the Keck Observatory Database of Ionized Absorbers toward QSOs (KODIAQ), funded through NASA ADAP grants NNX10AE84G and NNX16AF52G along with NSF award No. 1516777. Some of the measurements use observations collected at the European Southern Observatory. This research used resources of the National Energy Research Scientific Computing Center (NERSC), supported by the Office of Science of the U.S. Department of Energy under contract No. DE-AC02- 05CH11231. HST/COS data and metal line identifications from Danforth et al. (2016) were used in Chapter 4.



Eidesstattliche Versicherung

Bei der vorliegenden Dissertation handelt es sich um meine eigenständig erbrachte Leistung.

Ich habe nur die angegebenen Quellen und Hilfsmittel benutzt und mich keiner unzulässigen Hilfe Dritter bedient. Insbesondere habe ich wörtlich oder sinngemäß aus anderen Werken übernommene Inhalte als solche kenntlich gemacht.

Die Arbeit oder Teile davon habe ich bislang nicht an einer Hochschule des In- oder Auslands als Bestandteil einer Prüfungs- oder Qualifikationsleistung vorgelegt.

Die Richtigkeit der vorstehenden Erklärungen bestätige ich.

Die Bedeutung der eidesstattlichen Versicherung und die strafrechtlichen Folgen einer unrichtigen oder unvollständigen eidesstattlichen Versicherung sind mir bekannt.

Ich versichere an Eides statt, dass ich nach bestem Wissen die reine Wahrheit erklärt und nichts verschwiegen habe.

....	....
Ort und Datum	Unterschrift