Dissertation

submitted to the
Combined Faculties of the Natural Sciences and Mathematics
of the Ruperto-Carola-University of Heidelberg, Germany,
for the degree of
Doctor of Natural Sciences

Put forward by
Anna-Christina Eilers
born in San Francisco, United States of America

Oral examination: May 3$^{\text{rd}}$, 2019

Unravelling 13 Billion Years of Cosmic History with Spectroscopic Studies: From the Milky Way to the Epoch of Reionization

Anna-Christina Eilers

Referees: **Prof. Dr. Joseph F. Hennawi**

Priv.-Doz. Dr. Christian Fendt

Abstract

In this thesis we trace the cosmic history from the Epoch of Reionization to the local universe by means of several spectroscopic studies. In the first part, we analyze 34 quasar spectra at $5.8 \lesssim z_{\mathrm{em}} \lesssim 6.5$ and measure the redshift evolution of the opacity of the intergalactic medium (IGM) within the Lyα as well as the Lyβ forest to set new constraints on the timing and morphology of the reionization epoch. We find evidence for an extended reionization process down to $z \sim 5$, and, while the observed scatter in the Lyα forest optical depth can be well reproduced by current state-of-the-art simulations including spatial fluctuations in the temperature field or the ultraviolet background, we find a strong mismatch between simulations and observations in the Lyβ forest opacity, suggesting an inversion of the thermal state of the post-reionization IGM.

We also measure the sizes of the quasars' proximity zones, which are regions of enhanced ionization in the vicinity of the quasars, ionized by their own radiation. We find a dependency of the proximity zone sizes to the quasars' lifetime, which presents a novel method to estimate the lifetime of individual quasars, providing unprecedented constraints on the formation and growth of supermassive black holes in the early universe. We discover three quasars with very short lifetimes, i.e. $t_{\mathrm{Q}} \sim 10^4 - 10^5$ yr, that pose significant challenges to all current black hole formation theories.

In the second part of this thesis we explore the structure and dynamics of the Milky Way. We develop a new data-driven model to determine precise parallaxes by combining multi-band photometry and spectroscopy to make global kinematic maps of our Galaxy from $\gtrsim 45,000$ luminous red giant stars with only $\lesssim 10\%$ parallax uncertainties. Our map extends to Galactocentric distances of 25 kpc, well beyond the reach of parallax estimates by the *Gaia* mission. Making use of these new spectrophotometric parallaxes, we determine the most precise measurement to date of the circular velocity curve of the Milky Way over a wide range of Galactocentric distances. Based on Jeans modeling in an axisymmetric gravitational potential we find that the velocity curve is gently but significantly declining.

Zusammenfassung

In dieser Doktorarbeit werden wir die Geschichte unseres Universums, angefangen bei der Reionisationsepoche bis zum lokalen Universum, anhand von spektroskopischen Daten untersuchen. Im ersten Teil analysieren wir die Spektren von 34 Quasaren bei Rotverschiebungen $5.8 \lesssim z_{\mathrm{em}} \lesssim 6.5$ und messen die Entwicklung der Opazität des intergalaktischen Mediums (IGM) im Lyα und Lyβ Wald, um neue Randbedingungen zum zeitlichen Ablauf, sowie zur Morphologie des Reionisationsprozesses zu erlangen. Wir finden Hinweise auf eine ausgedehnte Reionisationsepoche bis $z \sim 5$. Obwohl die Verteilung der gemessenen Lyα optischen Tiefen gut von modernen hydrodynamischen Simulationen, die räumliche Fluktuationen in der ultravioletten Hintergrundstrahlung oder in der Temperaturverteilung des intergalaktischen Gases erfassen, vorausgesagt werden kann, finden wir eine starke Diskrepanz zwischen den Simulationen und den beobachteten optischen Tiefen im Lyβ Wald, was auf eine Inversion des thermischen Zustandes des IGM nach der Reionisationsepoche deutet.

Des Weiteren messen wir die Größe der Nahzonen der Quasare, welche die Regionen in deren Umgebung beschreiben, in denen die Ionisierung des intergalaktischen Gases höher ist aufgrund der eigenen Strahlung der Quasare. Wir entdecken einen Zusammenhang zwischen der Größe der Nahzonen und dem Alter der Quasare, was eine neue Methode zur Altersbestimmung einzelner Quasare darstellt, mit der wir noch nie zuvor möglich gewesene Bedingungen für die Entstehung und das Wachstum supermassereicher schwarzer Löcher im frühen Universum ableiten können. Drei der Quasare, die wir untersuchen, zeigen ungewöhnlich kurze Lebensdauern auf von nur $\sim 10^4 - 10^5$ Jahren, die sämtliche derzeitige Theorien, die das Wachstum schwarzer Löcher zu erklären versuchen, in Frage stellen.

Im zweiten Teil dieser Arbeit wenden wir uns der Struktur und Dynamik der Milchstraße zu. Wir entwicklen ein neues Daten gestüztes Model, um präzise Messwerte für Parallaxen von $\gtrsim 45,0000$ roten Riesensternen anhand derer spektroskopischen und photometrischen Daten zu ermitteln. Mithilfe dieser Parallaxen, die wir mit nur $\lesssim 10\%$ Ungenauigkeit bestimmen können, gelingt es uns, eine globale Karte unsere Galaxie zu erstellen, die sich über mehr als 25 kpc erstreckt, was weit jenseits der Distanz liegt, für die die *Gaia* Mission gute Parallaxen messen kann. Wir nutzen diese neuen spektrophotometrischen Parallaxen, um die Rotationskurve der Milchstraße über weite Entfernungen vom Galaktischen Zentrum aus präzise zu bestimmen. Basierend auf einem Jeans Model in einem achsensymmetrischen Gravitationspotential, erkennen wir, dass die Rotationskurve langsam, aber signifikant abfällt.

Für meine Familie.

Preface

In the few weeks while writing up this thesis a jigsaw puzzle appeared in the coffee area of MPIA. It had a thousand pieces and was depicting the Whirlpool Galaxy M51 and *a lot* of nearly dark sky around it, scattered with a few stars. No one knew who had put the puzzle there, it had just turned up after the Christmas break. Every time I was walking passed it I saw someone standing around it, trying to add another piece to the puzzle. The yet unplaced pieces got sorted by shape, by the number of stars, or by slight variations in the brightness of the background – and then they got re-ordered and re-shuffled again. Slowly, piece by piece, the image grew and evolved.

This thesis is in many ways very similar to the jigsaw puzzle. It is the contribution of a small piece to the big puzzle of our universe. And it is the product of much support, patience, and friendship of many people without whom this work would have not been possible.

In this thesis we will unravel 13 billion years of cosmic history by means of spectroscopic data from multiple observatories around the world. In the first part we will trace the early evolution of our universe by observations of distant quasars. The main questions I aim to address are *What do quasar spectra reveal about the timing and morphology of the Epoch of Reionization?* and *How long do quasars shine and what does this imply for the formation of supermassive black holes?* The second part focuses on the very local universe and studies the structure and dynamics of our home galaxy, the Milky Way. Based on a new data-driven model that makes use of spectroscopic and photometric properties of luminous giant stars, I want to know *What is the shape and slope of the circular velocity curve of the Milky Way?* The results of these studies are published in six scientific papers and are presented in the Chapters 8 to 14 in this thesis.

Images of broken light which dance before me like a million eyes,

That call me on and on across the universe.

– The Beatles, Across the Universe

Contents

Contents

Contents

List of Figures

List of Tables

Part I

Context and Theoretical Background

The Standard Model of Cosmology

One of the main challenges of observational cosmology today is to understand how the first luminous objects in the universe formed and evolved through cosmic time. In part II of this thesis, we will start with the early epochs of the universe and trace the cosmic history by means of distant quasars, which are the most luminous, non-transient sources known in the universe, powered by accretion onto a central supermassive black hole (SMBH). In part III of this thesis we will have a close look at the local universe and analyze the dynamics of our own galaxy, the Milky Way.

We will start with a brief review of cosmological concepts in the framework of our current standard model ΛCDM that we will need throughout this thesis. The main references for this chapter are Dodelson (2003) and Liddle (2003).

1.1 The Expansion of the Universe

Almost all galaxies appear to be moving away from us due to the expansion of the universe. The recession velocity $\vec{v}(t)$ at a time t is proportional to the distance $\vec{r}(t)$ of an object, i.e.

$$\vec{v}(t) = H_0 \vec{r}(t), \qquad (1.1)$$

which is known as the *Hubble-Lemaître law*[1], where H describes the Hubble parameter, and the index "0" indicates its present day value. This relation is not exact as it does not hold perfectly for nearby galaxies due to their peculiar motion, but it describes the average behaviour of galaxies extremely well. The observed light emitted by a distant object becomes redshifted, i.e. the wavelengths are being stretched by the expansion of the universe and its change therefore tells us, how

---

[1]previously known as *Hubble's law*, renamed by the International Astronomical Union (IAU) on October 26th, 2018 (https://www.iau.org/news/pressreleases/detail/iau1812/)

much the universe has expanded since the light began to travel. This cosmological redshift z is defined as

$$1 + z = \frac{\lambda_{\text{obs}}}{\lambda_{\text{em}}} = \frac{a(t_{\text{obs}})}{a(t_{\text{em}})}, \tag{1.2}$$

where λ_{em} and λ_{obs} are the emitted and observed wavelengths, respectively. The quantity $a(t)$ is known as the scale factor of the universe, which determines the universal expansion rate. It is a function of time alone and tells us how physical separations are growing with time. At the present epoch, which is usually the time of the observations, i.e. $t_{\text{obs}} = t_0$, the scale factor is set to unity, thus $a(t_{\text{obs}}) = 1$.

The behaviour of the scale factor and hence the dynamics of the universe are governed by the *Friedman equations*, which are solutions of Einstein's field equations for a homogeneous and isotropic universe:

$$H^2 = \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{\kappa c^2}{a^2} + \frac{\Lambda}{3} \tag{1.3}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + \frac{3p}{c^2}\right) + \frac{\Lambda}{3}, \tag{1.4}$$

where G is the gravitational constant, κ describes the curvature of the universe, c defines the speed of light, and Λ represents the cosmological constant, which accounts for the effect of dark energy that drives the accelerated expansion of our universe (Riess et al., 1998; Perlmutter et al., 1999). These equations connect both the pressure p and density ρ of the universe to the scale factor.

Assuming a flat geometry of our universe, i.e. $\kappa = 0$, which is in accordance with the latest results by the Planck Collaboration et al. (2018a), and setting $\Lambda = 0$, we can determine the critical density of our universe, i.e.

$$\rho_{\text{crit}}(t) = \frac{3H^2}{8\pi G}. \tag{1.5}$$

At the present time this results in

$$\rho_{\text{crit},0} = 2.78 h^{-1} \times 10^{11} M_\odot / (h^{-1}\text{Mpc})^3, \tag{1.6}$$

where $h = \frac{H_0}{100\,\text{km}\,\text{s}^{-1}\,\text{Mpc}^{-1}}$ is the dimensionless Hubble parameter. Since $10^{11} - 10^{12} M_\odot$ is the mass of a typical galaxy, and one Mpc the typical separation between them, the universe is close to the critical density.

We introduce the dimensionless density parameters for radiation Ω_r, matter Ω_m, and the cosmological constant Ω_Λ, i.e.

$$\Omega_r = \frac{\rho_{r,0}}{\rho_{\text{crit},0}}; \qquad \Omega_m = \frac{\rho_{m,0}}{\rho_{\text{crit},0}}; \qquad \Omega_\Lambda = \frac{\rho_\Lambda}{\rho_{\text{crit},0}} = \frac{\Lambda}{3H_0^2}, \tag{1.7}$$

and

$$\Omega_0 = \Omega_m + \Omega_r + \Omega_\Lambda, \tag{1.8}$$

where $\Omega_0 = 1$ describes a flat geometry of our universe, whereas $\Omega_0 > 1$ or $\Omega_0 < 1$ represent an open or closed universe, respectively. We can now reformulate the 1$^{\text{st}}$ Friedman equation to

$$H^2 = \left(\frac{\dot{a}}{a}\right)^2 = H_0^2 \left[\Omega_r a^{-4} + \Omega_m a^{-3} + \Omega_\Lambda + (1 - \Omega_0) a^{-2}\right], \tag{1.9}$$

which enables us to divide the evolution of the universe into three regimes: a radiation-dominated era at very early cosmic times ($t \lesssim 47,000\,\text{yr}$), followed by a matter-dominated regime ($47,000\,\text{yr} \lesssim t \lesssim 9.8\,\text{Gyr}$), and the current epoch, which is dominated by the cosmological constant Λ.

1.2 Cosmological Redshift and the Age of the Universe

In order to understand the formation of the first objects in the universe, it is essential to have an estimate of the age of the universe at different cosmological redshifts. Integrating Eqn. (1.9) provides an estimate of the age of the universe as a function of the scale factor, i.e.

$$t(a) = \frac{1}{H_0} \int_0^{a'} \left(\frac{\Omega_r}{a^2} + \frac{\Omega_m}{a} + (1 - \Omega_0) + \Omega_\Lambda a^2\right)^{-1/2} \mathrm{d}a. \tag{1.10}$$

When combining this result with Eqn. (1.2) we can determine the age of the universe at the time when the light of an object at redshift z had been emitted.

Assuming the most recent estimates of cosmological parameters for a universe with a flat ΛCDM cosmology, i.e. $H_0 = 67.4 \pm 0.5\,\text{km}\,\text{s}^{-1}\,\text{Mpc}^{-1}$, $\Omega_m = 0.315 \pm 0.007$, and $\Omega_\Lambda = 0.685 \pm 0.007$ (Planck Collaboration et al., 2018a), the age of the universe at redshift $z = 5$ (6, 7) is $t = 1.2$ Gyr (927 Myr, 759 Myr). At the present time, i.e. $z = 0$, the universe is approximately 13.8 Gyr old.

1.3 Distances

In cosmology there are many ways to specify the distance between two objects, because distances are constantly changing due to the expansion of the universe. Large distances to far-away objects are often determined in terms of their redshift, since redshifts usually are the only quantity related to distance that can be observed directly.

- The *proper distance* describes the distance between two objects measured along a geodesic at a specific, fixed moment in time. However, this distance will change with time due to the expansion of the universe.

- The *comoving distance* of two objects moving with the Hubble flow does not change in time, as it accounts for the expansion of the universe, i.e.

$$d_{\text{comoving}}(z) = c \int_0^z \frac{\mathrm{d}z'}{H(z')} \tag{1.11}$$

where $H(z')$ is given by Eqn. (1.9) and Eqn. (1.2).

The comoving distance and the proper distance are related by the scale factor, i.e.

$$d_{\text{proper}} = a(t) \cdot d_{\text{comoving}}, \tag{1.12}$$

and thus these distances are equal at the present day.

- The *angular-diameter distance* is defined by means of the object's physical extent l and its angular size θ when observed from Earth, i.e.

$$d_{\text{angular}}(z) = \frac{l}{\theta}, \tag{1.13}$$

which can be written in terms of the comoving distance as

$$d_{\text{angular}}(z) = \frac{d_{\text{comoving}}(z)}{1+z}, \tag{1.14}$$

when assuming a flat universe[2].

- The *luminosity distance* of an object can be expressed by its bolometric luminosity L and its observed flux F, as

$$d_{\text{luminosity}} = \sqrt{\frac{4\pi F}{L}}. \tag{1.15}$$

This quantity is relevant for measurements of distances to astronomical standard candles, such as Type Ia Supernovae for instance, for which we can infer their intrinsic absolute luminosity.

---

[2]In a universe with positive, i.e. $\kappa > 0$, or negative curvature, i.e. $\kappa < 0$, the enumerator in Eqn. (1.14) changes to $1/\sqrt{\kappa}\sin(\sqrt{\kappa}\cdot d_{\text{comoving}})$ or $1/\sqrt{-\kappa}\sin(\sqrt{-\kappa}\cdot d_{\text{comoving}})$, respectively.

1.4   A Brief History of the Universe

The history of our universe began approximately 13.8 billion years ago with the *Big Bang* (see Fig. 1.1). At this point in time the universe was in an extremely dense, hot, and highly energetic state, from which it started to expand. It is assumed that within the first 10^{-37} seconds, the universe experienced a very rapid, exponential growth phase, known as *cosmic inflation* (Guth, 1981). During this time density fluctuations caused by the uncertainty principle were amplified, setting the seeds that would later form the large-scale structure of the universe. After inflation the universe continued to expand but at a much slower pace.

The universe was filled with a hot, opaque plasma of baryons, electrons, and photons. About $10 - 20$ minutes after the Big Bang the temperature and pressure of the universe allowed for nuclear fusion to occur, enabling the formation of the first atomic nuclei, mainly hydrogen and helium with small traces of lithium, a process known as *primordial nucleosynthesis* (Alpher et al., 1948).

At redshift $z \approx 1100$, roughly $377,000$ years after the Big Bang, the universe had cooled down enough such that the photons could decouple from the baryons, and the atomic nuclei and electrons could combine to form neutral atoms. During this era known as *recombination*, the universe became transparent, and the decoupled photons, which we can still observe today as the *Cosmic Microwave Background* (Penzias & Wilson, 1965), could travel freely through the universe.

Following recombination, the cosmic *dark ages* began, during which there were no sources of visible light yet. Dark matter had collapsed into halo-like structures driven by its own gravitational attraction. Ordinary, baryonic matter was pulled into those potential wells, ultimately collapsing into clouds of predominantly hydrogen, from which the first stars and galaxies formed (see Chapter 2 for details). This first generation of stars, the yet unobserved, so-called population III stars, is believed to have formed out of the very pristine primordial gas. All heavier elements that we observe in the present day universe were formed by nuclear fusion subsequently in the cores of stars.

The formation of the first stars, galaxies, and quasars and the release of their ultraviolet (UV) radiation ushered in a new cosmic era, the *Epoch of Reionization*. During this early evolutionary phase of our universe, the cosmic dark ages ended, and the gas in the universe, the intergalactic medium (IGM), transitioned from a neutral state into the mostly ionized medium that we observe today. The details of the reionization process not only reflect the nature of these primordial objects, but also the formation of large-scale structure and are therefore a subject of major
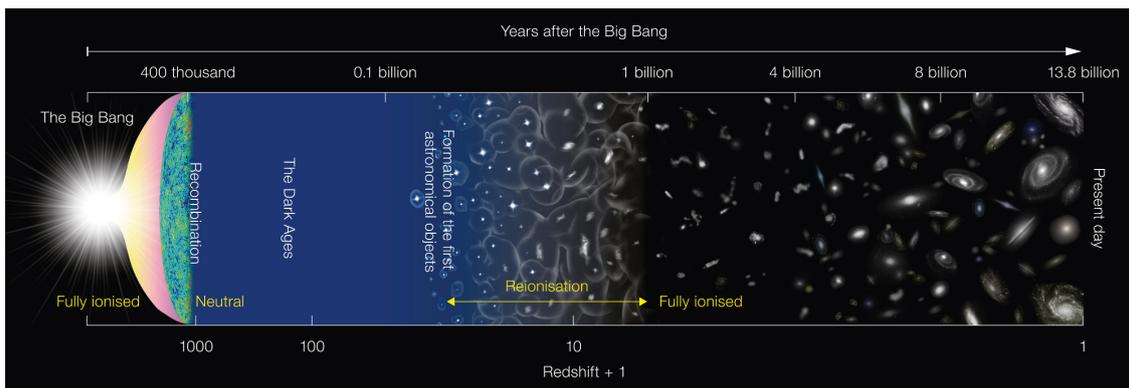
Figure 1.1 Schematic representation of the evolution of the universe covering 13.8 billion years of cosmic history, from the Big Bang to the present day (Figure Credit: NAOJ).

interest. Determining when and how the Epoch of Reionization proceeded is one of the primary goals of observational cosmology today.

Despite much progress in the last decade, there are still crucial yet unanswered questions regarding the timing, duration, and morphology of the reionization epoch. For instance, the beginning of the reionization process is still highly uncertain, but there are some observational constraints (see e.g. Fan et al., 2006a, for a review). The most recent measurement of the Thomson optical depth $\tau = 0.054 \pm 0.007$ (Planck Collaboration et al., 2018a) suggests the mid-point of reionization was at $z_{\mathrm{reion}} = 7.7 \pm 0.7$. Further constraints on the end stages of the reionization epoch can be inferred from the measurements of the redshift evolution of the opacity of the IGM, which we will address in Chapter 9 and 10.

After reionization was complete, the structure in the universe continued to grow. At $z \sim 2$ the star formation and quasar activity were at their peak, and the most growth is believed to occur (e.g. Madau & Dickinson, 2014). The present day universe at $z = 0$ shows a huge variety of cosmic structures, aligned in sheets and filaments, forming the cosmic web. Among the various objects we find active, very luminous galaxies, so-called quasars – which are the focus of part II of this thesis –, as well as massive elliptical galaxies, and beautiful spiral galaxies, such as our home galaxy, the Milky Way – which we will study in more detail in part III of this thesis.

The Formation of Cosmic Structure

All structure that we observe in the universe today grew by gravitational instabilities (Jeans, 1902) and hierarchical merging out of the initial density perturbations introduced during cosmic inflation. The inflationary theory predicts that the primordial density perturbations follow a Gaussian random field, and the power spectrum of these initial density fluctuations, $P(k) \propto k^n$, is scale-invariant, i.e. $n \sim 1$ (Hawking, 1982; Guth & Pi, 1982; Bardeen et al., 1983). The latest results from the Planck Collaboration, $n = 0.9649 \pm 0.0042$ (Planck Collaboration et al., 2018b), confirm this scale-invariance, although with small discrepancies from the expected value.

Observations of the Cosmic Microwave Background (CMB) show that at the time of recombination the universe was still very uniform with spatial fluctuations in the density field of only one part in 10^5 (e.g. Smoot et al., 1992). However, while baryonic perturbations were damped by photon pressure, so-called *Silk damping* (Silk, 1968), which made the fluctuations in the CMB appear more uniform, density perturbations in the dark matter could already grow before recombination. At recombination, the baryons decouple from the photons and condense into the gravitational potential wells set by dark matter, which are the building sites for the formation of the first stars and galaxies (White & Rees, 1978; Fall & Efstathiou, 1980).

Thereby gravitational instabilities of the small initial density perturbations collapse first along one axis, forming sheets, which are also called *Zel'dovich pancakes* (Zel'dovich, 1970), and then successively collapse along the other axes, building up filaments and finally dark matter halos. The large-scale structure in the universe that we observe today (Geller & Huchra, 1989; Colless et al., 2001) consists
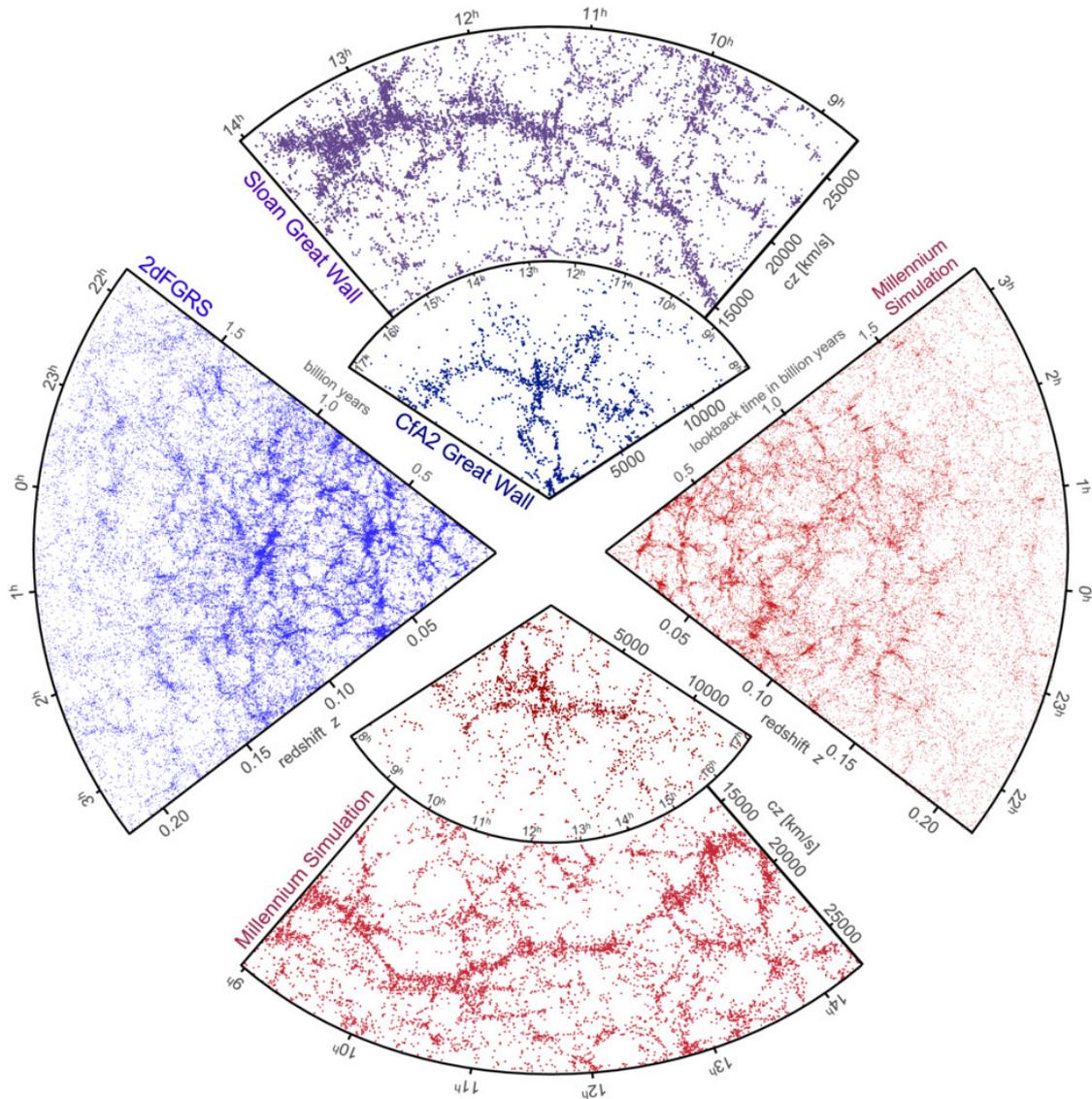
Figure 2.1 The observed (blue) and simulated (red) distribution of galaxies in the universe, which align along sheets and filaments forming the cosmic web (Figure Credit: Millennium Simulation, Springel et al. (2006)).

of galaxies and galaxy clusters within the most massive dark matter halos, connected by sheets and filaments that separate vast, nearly empty voids, forming the so-called *cosmic web*. Various numerical simulations have accurately reproduced the predicted and observed large-scale structure (see Fig. 2.1) since their dynamics are governed by gravity only (Gott et al., 2005; Springel et al., 2006; Frenk & White, 2012), and have thus established ΛCDM as the current standard model for cosmology and structure formation.
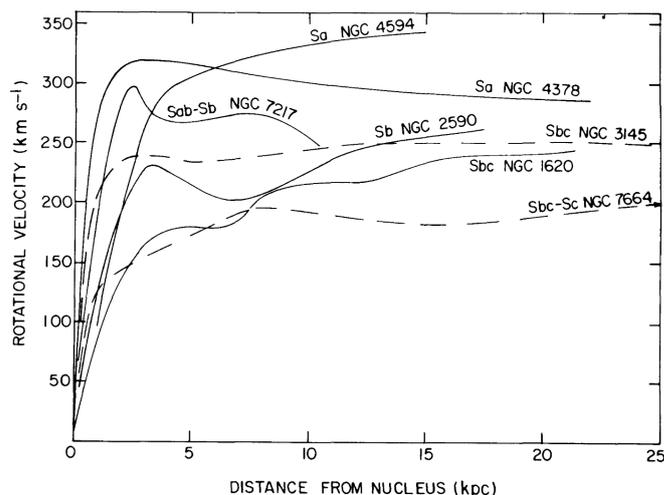
Figure 2.2 Radial profiles of the rotational velocities of various disk galaxies measured by Rubin et al. (1978). The velocities stay fairly constant with radius, instead of showing the originally expected declining Keplerian profile, which indicates the presence of a dark matter halo around the galaxies.

2.1 The Nature of Dark Matter

Dark matter plays a crucial role in structure and galaxy formation. Observations by Zwicky (1933), who analyzed the motion of stars in galaxies within the Coma cluster, presented the first indication that there could be a matter component present other than the visible stars, gas, and dust. Using the virial theorem, i.e. $K = -\frac{1}{2}V$, where K is the kinetic and V the gravitational energy, he concluded that the density of the Coma cluster would have to be ~ 400 times larger than suggested by the luminous matter.

A few decades later, Rubin et al. (1978, 1980) measured the rotation curve of several galaxies and found that the rotational velocity of stars v_c stays approximately constant with radius R, instead of showing a declining radial profile as expected from Keplerian motion, i.e.

$$v_\mathrm{c}(R) = \sqrt{\frac{G M(R)}{R}} \sim \text{const.,} \qquad (2.1)$$

where G denotes the gravitational constant and $M(R)$ the enclosed mass of the galaxy. This flatness of the rotation curves indicates that a matter component with a much larger mass than observed by the luminous matter is needed to keep the stars on their orbit. These results imply that galaxies are embedded in large spheroidal dark matter halos, which provides the missing mass component to explain the observations.

Today we believe that dark matter is cold, i.e. with a short free-streaming length, and has a small scattering cross-section, i.e. collisionless, interacting only via its gravitational pull. According to the current cosmological standard model ΛCDM, it consists of massive, non-baryonic particles. The homogeneity of the CMB provides evidence for its non-baryonic nature, because there would not have been

enough time to form the observed structure in the universe today, if the density perturbations at the time of recombination were as small as suggested by measurements of the CMB. Stronger fluctuations in the dark matter density field are needed to form potential wells prior to recombination, in which the baryonic matter can fall to form stars and galaxies once decoupled from the photons. The currently best candidate for dark matter particles are *weakly interacting massive particles* (WIMPs), which only interact via gravity and a weak force.

The behaviour of dark matter on large scales, where gravity dominates is well understood (e.g. Gott et al., 2005; Springel et al., 2006; Frenk & White, 2012), but on small scales there are still some open questions. Three main problems have been identified within the framework of ΛCDM due to a mismatch between cosmological N-body simulations and observations on small scales: the so-called *missing-satellite problem*, denoting a higher density of dark matter subhalos predicted in simulations than observed (e.g. Klypin et al., 1999), the *core-cusp problem*, designating the discrepancy between the flat density profiles of dwarf galaxies and the cuspy profile predicted by simulations (e.g. Moore et al., 1999), and the *too-big-to-fail problem*, arising from the ΛCDM prediction of satellites that are too massive and too dense, compared to those observed (e.g. Boylan-Kolchin et al., 2011, 2012).

Although it has been argued that most, if not all, of these discrepancies between observations and simulations can be resolved when including baryonic physics into the simulations (see Bullock & Boylan-Kolchin, 2017, for a recent review), these "problems" and the fact that the dark matter particle (if it exists) is yet to be found, have been motivation for various alternative dark matter models and modifications to ΛCDM. Among the most prominent alternatives are (i) *warm dark matter*, whose higher thermal velocity predicts the production of fewer dark matter subhalos, or (ii) *massive compact halo objects* (MACHOs), i.e. objects that emit very little or no light, including neutron stars, black holes, and brown and white dwarfs. However, this would suggest that dark matter would consist of baryonic material, and thus causing tension with the observations of the homogeneity of the CMB, as mentioned before. Another alternative are (iii) models of *modified gravity* that modify Newton's law of gravity in the limit of very low accelerations. This latter scenario has the advantage that it does not postulate the existence of any dark matter particles, however, it fails to explain other astronomical observations, such as the mass distribution within the bullet cluster for instance (Clowe et al., 2006).
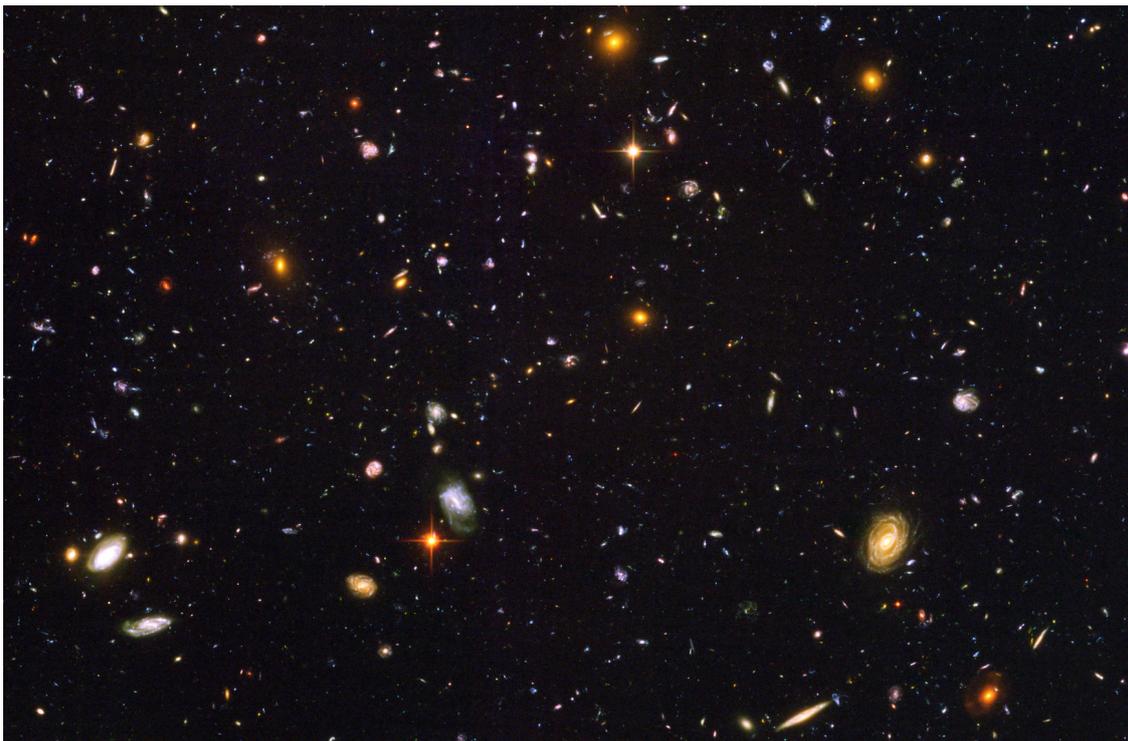
Figure 2.3 The Hubble Ultra Deep Field (HUDF) showing a patch of the sky scattered with galaxies of different morphologies, such as spiral galaxies, elliptical galaxies, as well as lenticular and irregular galaxies (Figure Credit: NASA, ESA, and S. Beckwith (STScI) and the HUDF Team).

2.2 Galaxy Formation and Morphology

Dark matter halos grow via hierarchical merging, settling eventually into virialized spheroidal structures due to their dissipationless nature, which follow a universal density profile that bends from a steep outer profile to a milder slope for the inner cusp, with a commonly adapted Navarro-Frenk-White (NFW)-profile (Navarro et al., 1997), i.e.:

$$\rho(R) = \frac{\rho_0}{\frac{R}{R_s}\left(1+\frac{R}{R_s}\right)^2},\tag{2.2}$$

where R_s is the scale length of the halo. For details on the involved processes we refer the reader to Mo et al. (2010), as well as Loeb & Furlanetto (2013).

The baryonic gas condenses into the potential wells and becomes shock heated to the virial temperature of the dark matter halo, at which it is supported against gravity by its thermal pressure. In order to form the first stars the gas has to cool down via radiative cooling of atomic and molecular hydrogen, as well as heavier elements in the subsequent generations of stars. The giant molecular gas clouds eventually collapse under their own gravity and fragment to form stars (see Naab

& Ostriker, 2017, for a recent review). In order to maintain star formation gas is replenished and accreted onto the dark matter halos via cold gas streams, eventually settling into a rotationally supported disk due to conservation of angular momentum, completing the formation of a disk galaxy (Fall & Efstathiou, 1980; Mo et al., 1998; Dekel et al., 2009).

To form elliptical galaxies it has been argued that a merger of two equal-mass galaxies can remove angular momentum and produce dispersion dominated spheroidal galaxies (e.g. Toomre, 1977; Hernquist, 1992). However, recent modeling of gas-rich mergers has shown that a disk can be re-established in post-merger galaxies due to subsequent gas accretion (Robertson et al., 2006; Hopkins et al., 2009). Thus additional mechanisms, such as stellar feedback, feedback from the active galactic nucleus (AGN), as well as quenching, need to be invoked to remove enough gas from the galaxies and prevent further gas accretion, in order to form quiescent elliptical galaxies (see Somerville & Davé, 2015, for a recent review).

The Milky Way

Our home galaxy, the Milky Way, is a fairly typical large disk galaxy with multiple spiral arms. From our point of view, however, the Milky Way appears as a thin milky white band stretched across the entire sky (see Fig. 3.1). This milky appearance gave our galaxy its name and resulted in the use of the word "galaxy", which has its origins in the Greek words $\gamma\alpha\lambda\alpha\xi\acute{\iota}\alpha\zeta$ $\kappa\acute{\upsilon}\kappa\lambda o\zeta$ (*galaxías kýklos*), meaning "milky circle".

3.1 Galactic Components

The main stellar components of our Galaxy are its disk, threaded by spiral arms, an elongated bar, and a central bulge, embedded in a stellar halo. However, most of the gas and stars in the Milky Way lie within its disk. Our Sun is located approximately 25 pc above the mid-plane of the stellar disk (Jurić et al., 2008) and ~ 8.122 kpc (Gravity Collaboration et al., 2018) away from the Galactic center, around which it rotates with a velocity of $\sim 229\,\mathrm{km\,s^{-1}}$ (Eilers et al., 2019). In what follows we will give a brief overview of the components of the Milky Way, as well as their structure and substructure (see Bland-Hawthorn & Gerhard, 2016, for a recent review).

- **Stellar Disk.** Due to our position close to the mid-plane of the disk of the Milky Way, we see the Galaxy from an edge-on perspective. Stellar number count studies in the infra-red (IR) (e.g. Churchwell et al., 2009), as well as recent studies leveraging astrometric data from the *Gaia* satellite mission (Gaia Collaboration et al., 2018; Antoja et al., 2018; Trick et al., 2018; Mackereth et al., 2019) have revealed a wealth of kinematic and chemical substructure within the stellar disk, hinting towards a complex assembly and

Figure 3.1 The Milky Way as seen from Earth (Cerro Paranal, Chile). We see the thin plane of the stellar disk and the luminous Galactic bulge in the center, partially obscured by dust clouds (Figure Credit: A.-C. Eilers).

evolution of the Milky Way. Considerable effort has also been invested in uncovering the Galactic spiral arms within the stellar disk, however, the exact number of spiral arms and their precise location remain largely unknown to date (e.g. Antoja et al., 2016; Rezaei Kh. et al., 2018).

The Milky Way disk has $\sim 10^{10} - 10^{11}$ stars and a scale length of $R_{\mathrm{s,disk}} \sim 2 - 3$ kpc (Bland-Hawthorn & Gerhard, 2016). Historically, the stellar disk has been divided into two components, a "thick" and "thin" disk. However, this definition based on the spatial distribution of stars has turned out to be misleading and a distinction in the stellar chemistry space seems to be more appropriate. Due to the observed bimodality in the chemical parameters when looking at the [α/Fe] vs. [Fe/H] abundances, it is now common to refer to a "low-α" and a "high-α" component of the disk (e.g. Haywood et al., 2013; Nidever et al., 2014; Bensby et al., 2014). The α-enriched stars formed earlier with a faster enrichment history, whereas the low-α population contains younger stars.

- **Bar.** The Milky Way exhibits an elongated bar in its center extending out to $R = 5.0 \pm 0.2$ kpc (Wegg et al., 2015). Its orientation with respect to the sightline between the Sun and the Galactic center is at an angle of $28° - 33°$ (Wegg et al., 2015), and it rotates as a solid body with a pattern speed of

$\Omega_b = 43 \pm 9 \, \text{km}\,\text{s}^{-1}\,\text{kpc}^{-1}$ (Bland-Hawthorn & Gerhard, 2016). The low scale height of the bar, $h_z \approx 180$ pc (Wegg et al., 2015), suggests that the bar is a feature of the stellar disk, possibly driven by disk instabilities.

- **Bulge.** The central, three-dimensional structure of the Galactic bar is a box/peanut shaped bulge (Binney et al., 1997; McWilliam & Zoccali, 2010; Nataf et al., 2010; Wegg & Gerhard, 2013) with an X-shape structure that can be observed in direct imaging (e.g. Ness & Lang, 2016). This barred bulge shows cylindrical rotation, and consists out of mostly old and metal-enriched stars (Ortolani et al., 1995) on boxy orbits in the triaxial potential of the bar.

  In the central part of the bulge we observe a nuclear star cluster (e.g. Becklin & Neugebauer, 1968; Eckart et al., 1995). By carefully tracking the motion of stars within the cluster we can infer the presence of a SMBH lurking in the center of our Galaxy (e.g. Ghez et al., 1998; Schödel et al., 2009). It is referred to as Sgr A*, with a mass of $M_{\text{BH}} = (4.100 \pm 0.034) \times 10^6 \, M_\odot$ (Gravity Collaboration et al., 2018). Our Galaxy shows no nuclear activity at the present day, although it has been argued that it has been active in the past (e.g. Zubovas et al., 2011).

- **Halo.** The stellar disk is embedded in an approximately spherical halo containing dark matter, stars, and gas. The stellar halo contains a population of old, high-velocity, and metal-poor stars that make up roughly 1% of all stars in the Milky Way (Bland-Hawthorn & Gerhard, 2016). Modern stellar surveys (see § 3.4) reveal an abundance of substructure in the stellar halo, such as tidal streams and shells, provoked by the accretion and disruption of satellite dwarf galaxies (e.g. Schlaufman et al., 2009; Belokurov, 2013; Helmi et al., 2018). Due to the long dynamical timescales in the halo this substructure provides a fossil record of the Milky Way's accretion history.

3.2 The Circular Velocity Curve

The circular velocity curve of the Milky Way $v_c(R)$ is the Galactocentric rest-frame velocity with which test particles would move on circular orbits at radius R from the Galactic center, in an axisymmetric gravitational potential Φ of a disk galaxy, such as the Milky Way, i.e.

$$v_c^2(R) = R \frac{\partial \Phi}{\partial R}\bigg|_{z \approx 0},$$
(3.1)

where z denotes the height above the Galactic plane. This $v_c(R)$ and in particular its value at the Sun's Galactocentric radius $R_\odot$, provide important constraints on

the mass distribution of our Galaxy and the local dark matter density. The latter is crucial for interpreting and analyzing any direct as well as indirect detection experiments of dark matter, whereas the shape of the rotation curve is a fundamental parameter for models of the Galactic disk. The local circular velocity at the Sun's location plays an important role when placing the Milky Way in a cosmological context and asking for instance, whether it falls onto the *Tully-Fisher relation*, which is a tight empirical relation between the total luminosity of spiral galaxies and their maximum rotation velocity (Tully & Fisher, 1977; Klypin et al., 2002).

Previously, the Sun's circular motion has been inferred by measuring its velocity with respect to an object at rest within our Galaxy, such as Sgr A* (Reid & Brunthaler, 2004). However, this method only determines the tangential velocity of the Sun, which differs from the circular velocity of the Galaxy at the location of the Sun by the Sun's so-called peculiar velocity. The method also requires knowledge of the Sun's Galactocentric distance, the precision of which improved significantly by the recent result from the Gravity Collaboration et al. 2018 (see also Bland-Hawthorn & Gerhard, 2016, for a review). An alternative way to determine the circular velocity at the Sun's radius, which does not require a precise estimate of the Sun's location with respect to the Galaxy, is the modelling of tidal streams in the Galactic gravitational potential, which has been performed with the GD-1 stream (Koposov et al., 2010) and Palomar 5 (Küpper et al., 2015).

For the inner region of the Galaxy the shape of the circular velocity curve has been inferred via the tangent-point method, based on the observed kinematics of the H I or CO emission of the interstellar medium, under the assumption of purely circular orbits of the gas (Gunn et al., 1979; Fich et al., 1989; Levine et al., 2008). The circular velocity in the outer part of the Galaxy, i.e. outside of the Galactocentric radius of the Sun, has been measured by means of line-of-sight velocities and distances using a variety of tracers, such as the thickness of the H I layer (Merrifield, 1992), spectrophotometric distances of H II regions combined with radial velocities of associated molecular clouds (Fich et al., 1989; Brand & Blitz, 1993), radial velocity measurements of planetary nebulae (Schneider & Terzian, 1983), classical cepheids (Pont et al., 1997) or RR Lyrae stars (Ablimit & Zhao, 2017; Wegg et al., 2018), blue horizontal branch stars in the halo (Xue et al., 2009; Kafle et al., 2012), red giant branch and red clump stars stars in the Galactic disk (Bovy et al., 2012; Huang et al., 2016), or masers in high mass star forming regions (Reid et al., 2014). However, such modeling requires a numerous and luminous tracer population and has significant limitations if the chosen tracers are either quite rare, e.g. classical cepheids, or not very luminous and hence not observable to large dis-

tances, e.g. red clump stars. Additionally, uncertainties in the distance estimates can lead to significant biases in the analysis of the circular velocity curve.

In Chapter 14 we will present the most precise measurement of $v_c(R)$ to date based on luminous red giant stars, for which we determine precise spectrophotometric distances with a data-driven approach that we introduce in Chapter 13.

3.3 Jeans Theorem of Galactic Dynamics

We will analyze the circular velocity curve of the Milky Way (Chapter 14) by means of a Jeans model, and hence we introduce the concept here. The aim of a Jeans model is to find the steady-state solution to stellar systems in equilibrium. Such models are the primary tool to compare observations and theory of galaxy dynamics.

Since the Milky Way has $\sim 10^{11}$ stars, it is more feasible to describe the ensemble of stars in terms of a probability distribution function, i.e. $f(\mathbf{x}, \mathbf{v}, t)$, such that $f(\mathbf{x}, \mathbf{v}, t) \, \mathrm{d}\mathbf{x} \, \mathrm{d}\mathbf{v}$ denotes the probability of finding a star at time t with phase-space coordinates $(\mathbf{x}, \mathbf{v})$ within the given range.

In a smooth and static external gravitational potential Φ, the probability density of a stellar system in equilibrium has to evolve continuously in phase-space. This assumption enables us to derive the *collisionless Boltzmann equation* (see § 4.1 in Binney & Tremaine, 2008), which describes the equilibria of collisionless systems, i.e.

$$0 = \frac{\mathrm{d}}{\mathrm{d}t} f(\mathbf{x}, \mathbf{v}, t) = \frac{\partial f}{\partial t} + \mathbf{v} \cdot \frac{\partial f}{\partial \mathbf{x}} - \frac{\partial \Phi}{\partial \mathbf{x}} \cdot \frac{\partial f}{\partial \mathbf{v}}, \tag{3.2}$$

where we applied Hamilton's equations with a Hamiltonian $H = \frac{1}{2}v^2 + \Phi(\mathbf{x}, t)$. An integral of motion is any function $I[\mathbf{x}(t), \mathbf{v}(t)]$ that is only dependent on the phase-space coordinates, and conserved along any orbit, i.e.

$$0 = \frac{\mathrm{d}}{\mathrm{d}t} I = \mathbf{v} \cdot \frac{\partial I}{\partial \mathbf{x}} - \frac{\partial \Phi}{\partial \mathbf{x}} \cdot \frac{\partial I}{\partial \mathbf{v}}. \tag{3.3}$$

A comparison of Eqn. (3.2) and Eqn. (3.3) reveals that any function of the integrals of motion, i.e. $f(I[\mathbf{x}(t), \mathbf{v}(t)])$, yields a steady-state solution, i.e. $\partial/\partial t = 0$, of the collisionless Boltzmann equation, which is known as the *Jeans theorem* (Jeans, 1915).

3.4 Overview of Stellar Surveys

In Chapter 13 we will use spectroscopic and photometric data from various stellar surveys, and develop a new technique to estimate parallaxes which will be relevant

for future surveys. In the following we will give a brief overview of these surveys in chronological order.

- The Two Micron All Sky Survey (**2MASS**) was a ground-based near infra-red (NIR) photometric survey, covering the entire sky in three bands, J, H, and K, between the years of 1997 and 2001 (Skrutskie et al., 2006).

- The Wide-field Infrared Survey Explorer (**WISE**) was a space mission that surveyed the sky in the IR between the years of 2009 and 2011 (Wright et al., 2010). It resulted in an all-sky photometric catalog in four IR wavelength bands ranging from $3.4\,\mu$m to $22\,\mu$m.

- The APO Galactic Evolution Experiment (**APOGEE**) is a IR spectroscopic survey that aims to obtain high-resolution, high signal-to-noise data of more than $100,000$ red giant stars across our Galaxy, covering the Galactic bulge, bar, disk, and halo (Majewski et al., 2017). Due to the high-resolution spectra precise radial velocities and element abundances of stars can be derived, providing insights into the dynamical structure and chemical enrichment history of our Galaxy.

- The main aim of the **Gaia** space mission, which is funded by the European Space Agency, is to create the largest, three-dimensional map of the Milky Way by measuring the precise position, brightness, and color of more than a billion stars in our Galaxy (Gaia Collaboration et al., 2016). During its five year mission it will conduct multi-epoch observations to allow accurate estimates of the stars' parallaxes and proper motions.

- **SDSS-V** will survey the entire sky in multiple epochs collecting optical and IR spectra of more than 6 million objects (Kollmeier et al., 2017). It will start collecting data in 2020 and comprises three programs: the *Local Volume Mapper*, which will observe the interstellar medium in local galaxies, the *Black Hole Mapper*, which among other goals intends to trace the growth of black holes across cosmic time, and the *Milky Way Mapper*, which is a multi-object spectroscopic survey that aims to create a densely sampled global Galactic map from 4 million optical and NIR stellar spectra. The derived stellar abundances and kinematic properties will enable new constraints on Milky Way formation scenarios.

Quasars: Beacons from the Distant Past

Models of quasar-galaxy co-evolution posit that every massive galaxy underwent a luminous quasar phase, producing the SMBHs found ubiquitously in the centers of all bulge-dominated galaxies in the local universe (e.g. Soltan, 1982; Kormendy & Richstone, 1995). These quasars are the most luminous, non-transient sources in the universe and can thus be observed at very early cosmic epochs. They are powered by gas accretion onto SMBHs located in the galactic nuclei, and emit copious amounts of radiation.

4.1   Spectral Energy Distribution

All AGN emit strongly over the whole electromagnetic spectrum, ranging from radio frequencies to X-rays. Their high luminosities arise due to very efficient radiation processes, where the rest-mass energy of accreted matter is being converted into electromagnetic radiation.

The study of black hole–quasar co-evolution requires a characterization of the quasar properties, such as their intrinsic luminosity, the covering factor of the nuclear obscuring medium, the stellar mass and star formation rate of the host galaxies, and the amount of re-processed emission by dust, which are encoded in the quasars' spectral energy distribution (SED), an example of which is shown in Fig. 4.1. The most prominent features of these SEDs are the "infrared bump" at $\sim 3 - 20\mu$m, and the so-called "big blue bump" in the optical and UV wavelength regime (Sanders et al., 1989; Shang et al., 2011). The continuous emission forming the big blue bump is believed to arise from the accretion disk around the SMBH at
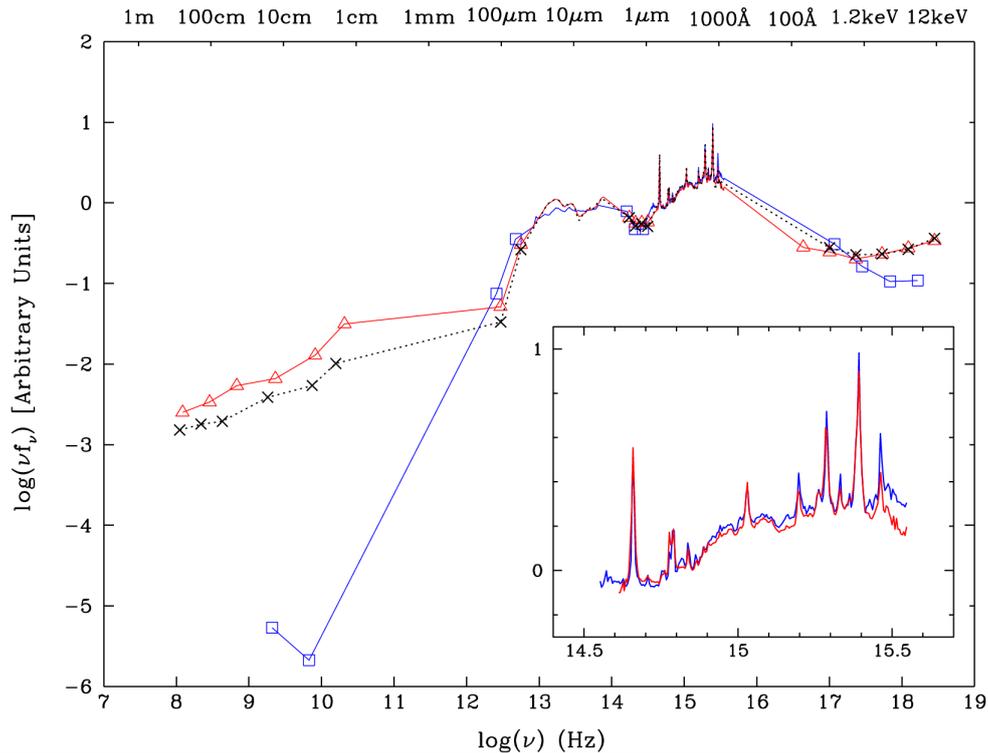
Figure 4.1 Composite SED for a sample of radio-loud (red) and radio-quiet (blue) quasars at $0.1 \lesssim z \lesssim 1.3$, as well as for a sample of radio-loud quasars at very low redshifts, i.e. $z < 0.5$, only (black). The inset shows the optical-UV region (Shang et al., 2011).

different radii. The infrared emission stems from dust that re-radiates a fraction of the optical and UV disk photons at infrared wavelengths.

Based on the presence of these large amounts of gas and dust surrounding the accretion disk, the so-called dusty torus, the quasar unification model has been developed (Antonucci, 1993; Urry & Padovani, 1995; Ramos Almeida & Ricci, 2017). This model explains the differences in the spectral characteristics of quasars as an effect of the viewing angle of the observer with respect to the dusty obscuring torus, which we will address in § 4.2 in more detail.

On top of the continuum emission from the quasar's accretion disk, the optical and UV spectrum typically shows broad emission lines, characteristic of moderately dense gas. The width of the lines arises due to the Doppler shift from clouds orbiting the SMBH. Making use of the virial theorem the velocity of the clouds allow an estimate of the mass of the central black hole – a technique we will apply in § 12.2.2 to calculate the mass of a black hole for one particular quasar of interest –, due to the gravitational force required to keep the clouds on their orbits and prevent them from escaping.

Additionally, most quasar SEDs also show narrow emission lines, which are expected to arise from highly ionized, more diffuse and slower-moving clouds of gas at larger radii. This Extended Narrow Line Region (ENLR) extends out to $15 - 20$ kpc and can be observed in Integral Field Unit (IFU) observations of high ionization lines, such as Hβ at 4861 Å and [O III] at 5007 Å (see e.g. Cracco et al., 2011; Liu et al., 2014).

In the radio-wavelength regime we observe two different types of AGN, so-called radio-loud and radio-quiet galaxies. Radio-loud galaxies account for about $10 - 20\%$ of the population and often exhibit bright radio lobes on either side of the galaxy. The emission is highly polarized synchrotron radiation that is believed to stem from electrons being accelerated along the magnetic field lines entangled within the radio jets.

It is important to note that the SEDs of quasars look similar independent of whether we observe the objects in the local or very distant universe (see Fig. 4.1), since the same physical processes govern the emission (e.g. Osmer, 2004; Fan, 2006; Shang et al., 2011).

4.2 Obscured and Unobscured Quasars

We mentioned previously that differences in the shape of the quasars' SEDs can be attributed to different viewing angles of the observer, as well as different covering factors of the nuclear obscuring medium (Fig. 4.2). This unification model of AGN (Antonucci, 1993; Urry & Padovani, 1995; Ramos Almeida & Ricci, 2017) suggests that the observer has a direct view on the inner region of quasars for so-called Type I quasars that exhibit broad emission lines, whereas for Type II quasars, the center and broad line region (BLR) are obscured by large amounts of gas and dust within the torus, and the observer only sees narrow emission lines from clouds in the ENLR. In these obscured quasars emission from the accretion disk cannot be directly detected.

However, this AGN unification scheme has been challenged by various studies (e.g. Sanders et al., 1988; Hopkins et al., 2008; Ishibashi & Fabian, 2016). Motivated by an observed connection between AGN activity and star formation bursts in a few well studied systems in the local universe, it has been argued that obscured quasars might be an evolutionary rather than a geometric phenomenon (see Hickox & Alexander, 2018, for a recent review).

The argument is that the AGN activity of a galaxy is being triggered by an interaction or merger, inserting a large supply of gas into the system which results
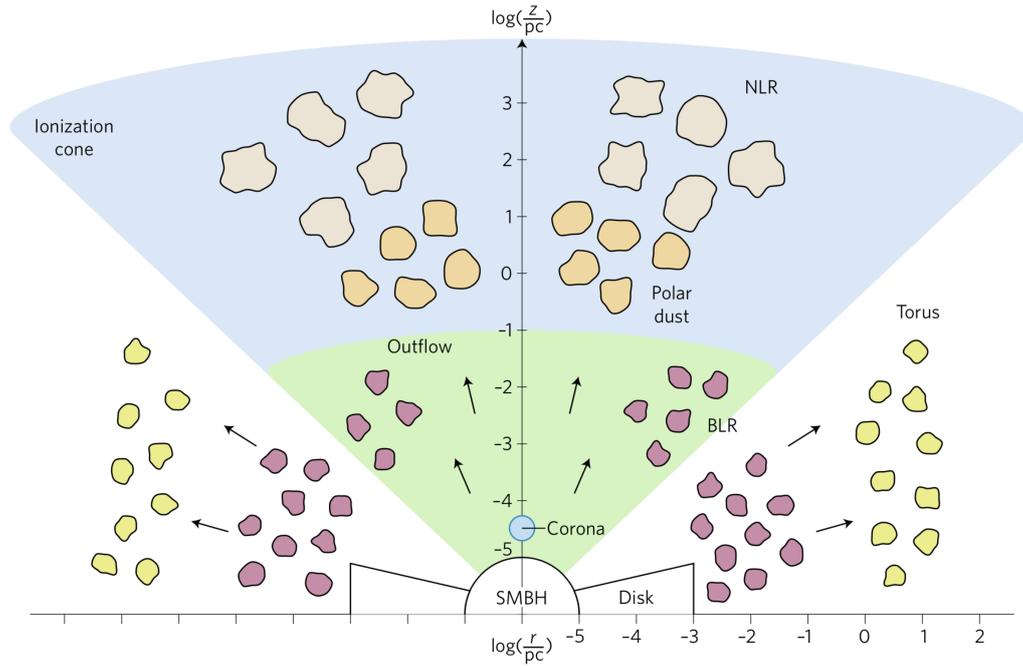
Figure 4.2 The unification model of AGN suggests that the observed differences in the SEDs of quasars can be attributed to different viewing angles of the observer and different covering fractions of the obscuring medium (Figure Credit: Ramos Almeida & Ricci (2017)).

in a burst in star formation (see Fig. 4.3). This scenario could explain the high star formation rate observed in quasar host galaxies, of up to $1000\,\mathrm{M_\odot\,yr^{-1}}$ (e.g. Walter et al., 2009). Due to the presence of large amounts of dust and gas the AGN is initially obscured. This early obscured quasar phase is followed by a blowout phase in which AGN feedback ejects gas from the galaxy, potentially quenching the starburst and revealing an unobscured quasar. In this model most of the growth of the SMBH is expected to occur in the initial dust-enshrouded phase. The subsequent AGN activity limits the black hole growth and prevents further star formation, explaining the observed tight relationship between SMBHs and host galaxies Hopkins et al. (2008); Hickox & Alexander (2018).

4.3 Quasars at High Redshift

In 1963 Caltech astronomer Maarten Schmidt identified the optical counterpart to a strong radio source 3C 273 (Schmidt, 1963). He was puzzled by the peculiar spectrum from this "star", before realizing that the bright emission lines he observed arise from strongly redshifted hydrogen gas ($z \approx 0.16$). Since the discovery of this first quasar in the 1960s, several hundred thousands of quasars have been

(c) Interaction/"Merger"
NGC 4676
- now within one halo, galaxies interact & lose angular momentum
- SFR starts to increase
- stellar winds dominate feedback
- rarely excite QSOs (only special orbits)

(d) Coalescence/(U)LIRG
NGC 6240
- galaxies coalesce: violent relaxation in core
- gas inflows to center: starburst & buried (X-ray) AGN
- starburst dominates luminosity/feedback, but, total stellar mass formed is small

(e) "Blowout"
IRAS Quasar Hosts
- BH grows rapidly: briefly dominates luminosity/feedback
- remaining dust/gas expelled
- get reddened (but not Type II) QSO: recent/ongoing SF in host high Eddington ratios merger signatures still visible

(f) Quasar
PG Quasar Hosts
- dust removed: now a "traditional" QSO
- host morphology difficult to observe: tidal features fade rapidly
- characteristically blue/young spheroid

(b) "Small Group"
M66 Group
- halo accretes similar-mass companion(s)
- can occur over a wide mass range
- M_{halo} still similar to before: dynamical friction merges the subhalos efficiently

(g) Decay/K+A
NGC 7252
- QSO luminosity fades rapidly
- tidal features visible only with very deep observations
- remnant reddens rapidly (E+A/K+A)
- "hot halo" from feedback
- sets up quasi-static cooling

(a) Isolated Disk
M81
- halo & disk grow, most stars formed
- secular growth builds bars & pseudobulges
- "Seyfert" fueling (AGN with $M_B > -23$)
- cannot redden to the red sequence

(h) "Dead" Elliptical
M59
- star formation terminated
- large BH/spheroid - efficient feedback
- halo grows to "large group" scales: mergers become inefficient
- growth by "dry" mergers

SFR [$M_\odot$ yr^{-1}]
$\log_{10}$ L$_{QSO}$ / L$_\odot$
Time (Relative to Merger) [Gyr]

a   b   c   d   e f g   h

Figure 4.3 Schematic description of a quasar model in which obscured and unobscured quasars are an evolutionary rather than a geometric phenomenon (Hopkins et al., 2008).

discovered (e.g. Pâris et al., 2018). However, it took almost half a century until the first quasars at high redshifts, i.e. $z \gtrsim 5.5$, were observed (Fan et al., 2000), which will be the focus of the first part of this thesis. In the past decade we have experienced a seven-fold increase in the number of known quasars at high redshifts (e.g. Venemans et al., 2015; Bañados et al., 2016; Mazzucchelli et al., 2017; Reed et al., 2017), which were discovered in deep wide-field optical and NIR surveys, such as the Sloan Digital Sky Survey (SDSS) and Pan-STARRS 1 (PS1). In total more than 350 quasars above $z \gtrsim 5.5$ are known to date. The current record holder, ULAS J1342 + 0928, was discovered by Bañados et al. (2018b) at a redshift of $z \approx 7.54$, which is less than 700 Myr after the Big Bang. These distant and very luminous quasars are rare objects with a number density of $\lesssim 1$ Gpc^{-3} (e.g. Wang et al., 2018).

High redshift quasars are of particular interest because they provide a unique window into the past history of our universe. These extreme objects represent an exciting opportunity to study the formation and evolution of the first sources in

the universe, as well as the early growth of SMBHs (see § 5). Due to the quasars' extreme luminosities they can be observed at very early cosmic times and used as beacons from the distant past, illuminating the space along the line-of-sight. This space in between quasars and galaxies is very vast, filled with extremely low density gas, the IGM. However, it is of enormous interest to study the properties of this low density gas, as well as its evolution with redshift, in order to understand the fuelling of galaxies and the physics of the reionization epoch (see § 6).

The Growth of Supermassive Black Holes

Observations of high redshift quasars indicate that they host SMBHs, i.e. $M_\mathrm{BH} \sim 10^9 - 10^{10} M_\odot$ (Mortlock et al., 2011; Venemans et al., 2013; De Rosa et al., 2014; Wu et al., 2015; Bañados et al., 2018b), in their center already at $z \gtrsim 6$, i.e. less than ~ 1 Gyr after the Big Bang. The formation and growth of these SMBHs at very early cosmic epochs is a crucial yet unanswered question in studies of black hole and galaxy evolution.

It has been argued that in order to grow the observed masses of SMBHs this early, very massive initial seeds are required, and, assuming accretion at the theoretical maximum rate, the so-called Eddington limit, that accretion must occur continuously for the entire age of the universe (Volonteri, 2010, 2012). These general considerations imply that the lifetime of quasars, the time in which matter is actively accreted onto the central SMBH, must be of the order of the Hubble time, i.e. ~ 1 Gyr at $z \sim 6$.

In the following we will have a closer look at the unknown parameters in studies of black hole growth: the origin and initial mass of the black hole seeds (§ 5.1), the radiative efficiency of the accretion (§ 5.2), and the lifetime of quasars (§ 5.3).

5.1   Black Hole Seeds

A black hole is the end product of the gravitational collapse of an object, such as a massive star. When a massive star runs out of fuel at the end of its life, it cannot support itself any longer against its own gravity and its core collapses to become a black hole. The remnants of the first generation of stars, the so-called population

III stars, are most commonly believed to provide the initial seeds for the SMBHs lurking in the heart of quasars (see Volonteri, 2010, for a review). The masses of these initial seeds are expected to be $M_{\mathrm{seed}} \sim 100\, M_\odot$ (Valiante et al., 2016). However, in order to grow the SMBHs observed in the earliest and most distant quasars from these stellar mass initial black holes, a constant mass accretion onto the black hole at the theoretical maximum rate (see § 5.2) throughout the whole age of the universe would be required, which is very difficult to maintain.

Another formation scenario of initial black hole seeds is based on the direct collapse of dense gas in gaseous, metal-free proto-galaxies (e.g. Lodato & Natarajan, 2006). Enriched halos cool very efficiently, which favors the fragmentation of the halos, whereas in metal-free conditions the formation of molecular hydrogen is inhibited and fragmentation is suppressed, providing the requirements for direct gas collapse. The direct collapse of massive gas clouds is expected to lead to the formation of very massive initial black hole seeds, i.e. $M_{\mathrm{seed}} \sim 10^4 - 10^6\, M_\odot$ (e.g. Visbal et al., 2014), which could explain the existence of SMBHs at early cosmic times. However, within the last few years simulations have shown that the conditions for these direct collapse black holes are very difficult to achieve (e.g. Johnson et al., 2014).

Other potential seeds for SMBHs form in stellar-dynamical processes. Since high redshift dark matter halos are likely already enriched with at least trace amounts of metals (Johnson et al., 2014), efficiently collapsing gas is more prone to form stars, instead of a direct collapse black hole. This early star formation episode can foster the formation of very compact nuclear star clusters, where stellar collisions within the star cluster can cause the build-up of a very massive star, leaving a remnant within the mass range $M_{\mathrm{seed}} \sim 10^2 - 10^4\, M_\odot$ (e.g. Omukai et al., 2008).

Fig. 5.1 illustrates the observed masses of the highest redshift quasars known to date and their potential seed masses, assuming continuous accretion at the theoretical maximum rate throughout the whole age of the universe. Under these (relatively unlikely) assumptions it is possible to grow the observed SMBHs from stellar remnant black hole seeds. Assuming a lower mass accretion rate or a shorter lifetime of the quasars would require more massive initial seeds, expected from direct collapse black holes or stellar-dynamical processes.

5.2 Radiative Efficiency of the Accretion Rate

If the infalling gas is gyrating around a central black hole, a centrifugal barrier develops, where the gas is held against gravity by rotation. Near the centrifugal
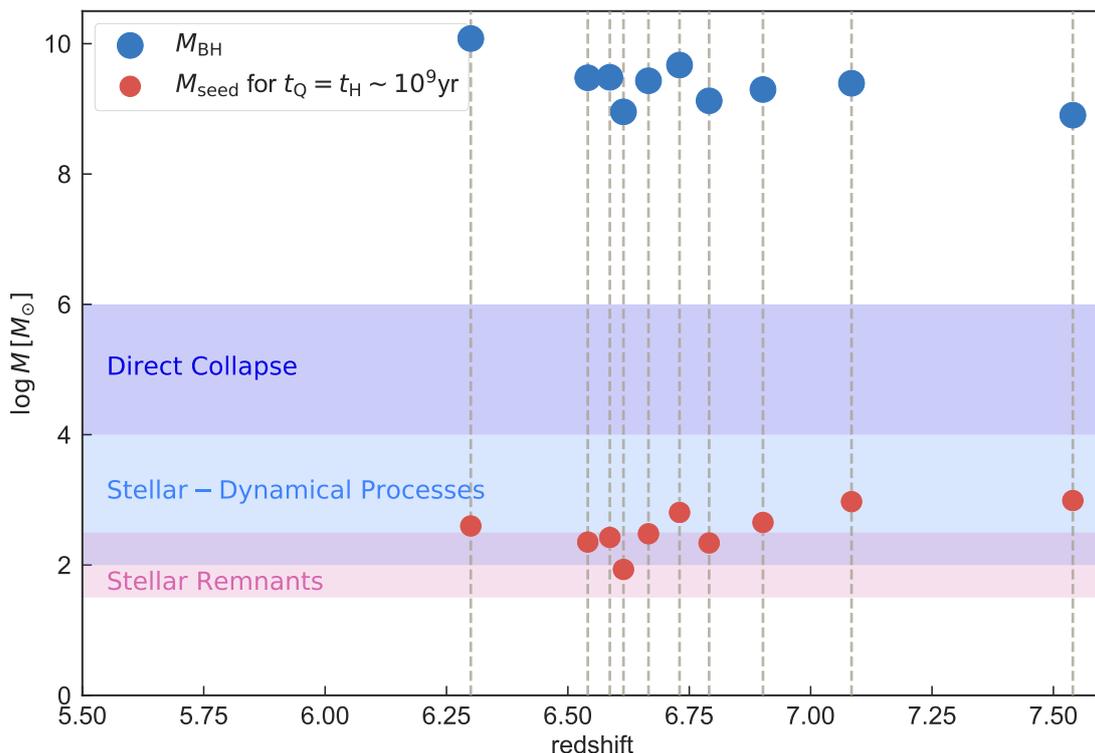
Figure 5.1 Observed masses of SMBHs (blue) of some of the highest redshift quasars known to date and their estimated black hole seed mass (red), assuming continuous accretion at the Eddington limit with a radiative efficiency of $\epsilon = 0.1$ throughout the whole Hubble time. The mass ranges of the initial black hole seeds produced by different formation scenarios are indicated by the shaded regions (adapted from Mazzucchelli et al., 2017).

barrier an accretion disk will form around the black hole. The accretion rate of matter onto the black hole is dictated by the rate at which angular momentum is being dissipated away through viscosity. As the gas settles onto the disk, its kinetic energy is dissipated into heat, and a thick and hot disk is established. However, if the cooling time of the gas is shorter than its viscous time, a thin accretion disk forms. This latter scenario is true for high gas inflow rates, which are believed to be present when fueling quasars (see Chapter 7 in Loeb & Furlanetto, 2013, for details).

There is a maximum luminosity, the so-called Eddington luminosity L_{Edd}, at which the environment of a black hole of mass M_{BH} can shine and still accrete matter, which can be derived by balancing the inward gravitational force of the black hole to the outward force of the radiation pressure, under the assumption of spherical symmetry, i.e.

$$\frac{G M_{\mathrm{BH}}(m_e + m_p)}{r^2} = \frac{L_{\mathrm{Edd}}}{4\pi r^2 c}\sigma_T,\tag{5.1}$$

where σ_T is the electron cross section for Thomson scattering, and m_e and m_p are the electron and proton mass, respectively. The resulting Eddington luminosity is

$$L_{\mathrm{Edd}} \approx 1.3 \times 10^{44} \left(\frac{M_{\mathrm{BH}}}{10^6 \, M_\odot} \right) \mathrm{erg\,s^{-1}}. \qquad (5.2)$$

The total luminosity L of the gas accreting onto a black hole can be described dependent on the radiative efficiency ϵ,

$$L = \epsilon \dot{M} c^2, \qquad (5.3)$$

where $\dot{M}$ is the total mass accretion rate. The black hole accretes the non-radiated component, i.e.

$$\dot{M}_{\mathrm{BH}} = (1 - \epsilon) \dot{M}. \qquad (5.4)$$

The growth of the black hole is then governed by

$$\dot{M}_{\mathrm{BH}} = \frac{M_{\mathrm{BH}}}{t_{\mathrm{S}}}, \qquad (5.5)$$

implying an exponential growth of its mass, i.e. $M_{\mathrm{BH}} \propto \exp(t/t_{\mathrm{S}})$ with an e-folding time scale, also known as "Salpeter timescale" (Salpeter, 1964), t_{S},

$$t_{\mathrm{S}} = 4.5 \times 10^7 \left(\frac{\epsilon}{0.1} \right) \left(\frac{L}{L_{\mathrm{Edd}}} \right) \mathrm{yr}, \qquad (5.6)$$

for a quasar radiating at a fraction L/L_{Edd} of the Eddington luminosity. Most studies suggest $\epsilon \approx 0.1$ and a luminosity of $L/L_{\mathrm{Edd}} \approx 1$, i.e. accretion at the Eddington limit (e.g. Yu & Tremaine, 2002), in agreement with predictions from models for a thin accretion disk. Black holes with lower radiative efficiencies are considered to have a *super-critical* mass accretion rate[1].

5.3 The Lifetime of Quasars

We will distinguish between several different timescales that govern the duration of quasar activity. The *duty cycle* t_{dc} refers to the total time that galaxies shine as active quasars integrated over the age of the universe. However, quasar activity could be episodic, and we refer to the duration of a single emission episode as the *episodic lifetime* t_{episodic}. The sizes of quasar proximity zones (see § 6.5) depend on the quasar *age* or the quasar's current *lifetime*[2]: If we denote by $t = 0$ the time at

---

[1]In the literature this term is often also referred to as "super-Eddington" accretion. This nomenclature, however, is misleading, since the Eddington limit only refers to the luminosity of the black hole environment rather than its accretion rate.

[2]Note that we will use both terms interchangeably throughout this thesis.

which light emitted by a quasar reaches our telescopes on Earth, then the quasar lifetime t_Q, is defined such that the quasar actually turned on at a time $-t_Q$ in the past. Strictly speaking the lifetime t_Q is a lower limit on the quasar episodic lifetime t_{episodic}, because the quasar episode might indeed continue, which we can only record on Earth if we could conduct observations in the future.

Measurements of quasar activity timescales have proven to be extremely challenging. At low redshifts, i.e. $z \sim 2-4$, the quasar duty cycle can be constrained by comparing the number density of quasars to their host dark matter halo abundance inferred from their clustering strength (Haiman & Hui, 2001; Martini & Weinberg, 2001; Martini, 2004; White et al., 2008). But to date this method has yielded only weak constraints on $t_{\text{dc}} \sim 10^6 - 10^9$ yr owing to uncertainties in how quasars populate dark matter halos (Shen et al., 2009; White et al., 2012; Conroy & White, 2013; Cen & Safarzadeh, 2015). Another possibility to infer the duty cycle of quasars has been presented by Soltan (1982), who pointed out that the luminosity function of quasars as a function of redshift reflects the gas accretion history of local remnant black holes, a statement known as the "Soltan argument". Following this argument Yu & Tremaine (2002) estimate the mean lifetime of luminous quasars from early-type galaxies found in SDSS to be $t_{\text{dc}} \sim 10^7 - 10^8$ yr. An upper limit on the duty cycle of quasars, $t_{\text{dc}} < 10^9$ yr, is set by the observed evolution of the quasar luminosity function, since the whole quasar population rises and falls over roughly this timescale (e.g. Osmer, 1998).

Various galaxy formation simulations that account for quasar feedback, as well as several studies of quasars in the local universe suggest that quasars exhibit variability in their activity on timescales ranging from days to $\sim 10^4$ yr, and hence they might accrete matter onto their SMBH in multiple short activity bursts (e.g. Ostriker et al., 2010; Ciotti et al., 2010; Novak et al., 2011; Choi et al., 2012). However, the constraints on the duty cycle of quasars do not give insights into the duration of the individual accretion episodes of the quasar activity. The clustering properties of a population of quasars emitting 1000 individual bursts each with a lifetime of $t_{\text{episodic}} \sim 10^5$ yr, would be indistinguishable from quasars with steady continuous emission for $t_{\text{episodic}} \sim 10^8$ yr.

So-called changing-look quasars for instance show variability in their photometric and spectral properties on timescales of days to decades, which has been attributed to changes in the accretion rate of the quasar (e.g. LaMassa et al., 2015; Runnoe et al., 2016; McElroy et al., 2016), although variable obscuration of the quasar cannot be excluded conclusively. Evidence for quasar variability on longer timescales of $\sim 10^4$ yr is based on the presence of relic emission nebulae around pre-

sumed quiescent galaxies (Schawinski et al., 2015; Sartori et al., 2016; Keel et al., 2017), or the existence of high-ionization absorption lines in the circumgalactic medium (CGM) of galaxies in the Cosmic Origins Spectrograph (COS)-halos survey (Oppenheimer et al., 2018). Quasar activity models with variability on short timescales, however, cannot explain the ubiquitous presence of large ionized nebulae around quasars at $z \sim 2$ extending out to several hundred kpc (Cantalupo et al., 2014; Hennawi et al., 2015; Cai et al., 2017, 2018), which require a continuous quasar activity for $t_{\mathrm{episodic}} \sim 10^5 - 10^6$ yr. Additionally, a lower limit on the individual quasar bursts of $t_{\mathrm{episodic}} \sim 3 \times 10^4$ yr is based on the argument that quasars need to maintain their ionizing luminosity long enough to explain the observed proximity zones in the Lyα forest (e.g. Bajtlik et al., 1988; Khrykin et al., 2016).

Additional uncertainty on the quasar lifetime estimate arises due to the unknown fraction of obscured quasars at high redshift. By analyzing the transverse proximity effect in the He II Lyα forest, i.e. the transmitted flux of background quasars in the presence of foreground quasars along the line-of-sight, Schmidt et al. (2018) found a bimodal distribution of quasar emission properties, where one population could be relatively old and unobscured, whereas the other could either be younger or highly obscured.

To conclude, there are still large uncertainties and disagreement in studies of the timescales of quasar activity to date. However, these timescales are a crucial ingredient for explaining the growth of SMBHs. Hence, in Chapter 11 of this thesis we will present a novel method to infer the lifetime of quasars.

The Physics of the Intergalactic Medium

The vast majority of matter in the universe lies outside of luminous galaxies in the IGM. This intergalactic gas traces the cosmic web, forming sheets and filaments in between nearly empty voids, that arise due to gravitational instability in a universe dominated by cold dark matter (see e.g. Cen et al., 1994; Miralda-Escudé et al., 1996; Davé et al., 1999). The very low density material consists of mostly pristine gas from the primordial nucleosynthesis. Hydrogen is the most abundant element, making up to approximately 93% of the atoms in the universe, and nearly all of the remaining matter is helium, with very small traces of lithium. All heavier elements were formed in the interiors of stars within galaxies.

Although dark matter dominates the mass budget of the IGM, we are primarily interested in the baryons, because the intergalactic gas ultimately provides the fuel for galaxy formation. However, most of the baryons remain in diffuse clouds that we can only observe when they intercept the light of a distant source. In the following we will look at the imprint of absorption lines from the intergalactic gas onto the spectra of distant quasars.

6.1 The Imprint of Lyman-α Absorption on Quasar Spectra

There are two important transitions of the hydrogen atoms that are widely used in astronomy: the 21-cm transition of hydrogen occurs between the two hyperfine states of the ground energy level $n = 1$ due to a spin flip of the electron. The second transition is the Lyα transition between the $n = 2$ and $n = 1$ energy levels, which
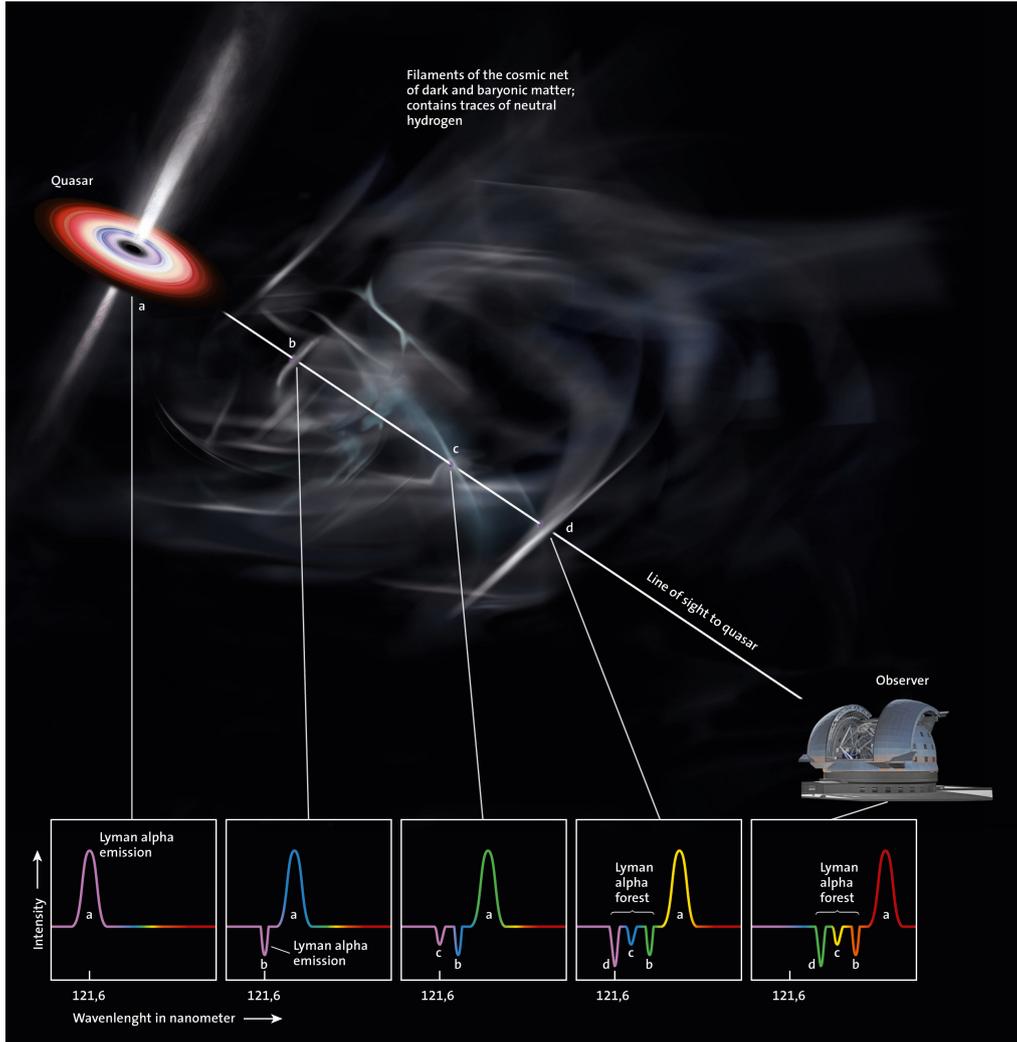
Figure 6.1 Schematic description of Lyα forest absorption lines observed in the spectra of distant quasars (Figure Credit: Sterne und Weltraum).

results in a Lyα photon of wavelength $\lambda_{\mathrm{Ly}\alpha} = 1216$ Å, and has been discovered experimentally in 1905 by physicist Theodore Lyman. This line has been used to probe the ionization and thermal state of the IGM in the spectra of quasars, galaxies, and gamma ray bursts, and will be employed in various analyses throughout this thesis.

While an emitted photon with wavelength $\lambda < \lambda_{\mathrm{Ly}\alpha}$ travels from a distant source through the universe towards us, it becomes redshifted due to the expansion of the universe and eventually redshifts into resonance with the Lyα absorption line (see Fig. 6.1). The photon can then be absorbed by traces of neutral hydrogen in the intergalactic gas, resulting in a characteristic absorption pattern in the spectrum of the source, which is called the Lyα forest (Lynds, 1971). The imprint of these redshifted Lyα forest absorption lines has become an important tool for constrain-

ing cosmology and the physics of the IGM at $2 \lesssim z \lesssim 6$ (see e.g. Croft et al., 1998; Weinberg et al., 2003; Zaldarriaga et al., 2003; McDonald et al., 2006; Meiksin, 2009; Becker et al., 2015a; Mortlock, 2016; McQuinn, 2016), because the observed absorption pattern traces the density and opacity fluctuations of the IGM along the filamentary structure of the cosmic web, enabling studies of the redshift evolution of the intergalactic gas throughout cosmic history (see Fig. 6.2).

In Chapter 10 we will explore a different tracer co-spatial with the Lyα forest, namely the Lyβ forest. Whereas the overly-sensitive Lyα transition saturates for neutral fractions $x_{\mathrm{HI}} \gtrsim 10^{-4}$, the ~ 5 times lower oscillator strength of the Lyβ transition, which is the transition between the energy levels $n = 3$ to $n = 1$ in the hydrogen atom, makes it sensitive to gas with a higher neutral fraction providing more stringent constraints on the IGM ionization state (Lidz et al., 2002; Songaila, 2004; Fan et al., 2006c). Furthermore, we know that for a gas in photoionization equilibrium the neutral gas fraction is proportional to the gas density, see Eqn. (6.2), and since the temperature of the intergalactic gas also correlates with its density, as shown in Eqn. (6.1), the ratio of Lyα to Lyβ forest opacity is sensitive to the thermal state of the IGM (Oh & Furlanetto, 2005).

6.2 The Thermal State of the Intergalactic Medium

An important characteristic of the fluctuating IGM is a tight relationship between the temperature T and the density ρ of the cosmic gas (Hui & Gnedin, 1997; McQuinn et al., 2009; Furlanetto & Oh, 2009; McQuinn & Upton Sanderbeck, 2016). This "equation of state" is controlled by the interplay of several mechanisms: photoionization heating by the ultraviolet background (UVB) radiation, inverse Compton cooling, and adiabatic cooling due to the expansion of the universe. Assuming a power-law relationship for the temperature-density relation we can express the equation of state as

$$T = T_0 \left(\frac{\rho}{\rho_0} \right)^{\gamma - 1}, \tag{6.1}$$

where T_0 denotes the temperature at average density ρ_0. A slope with $\gamma < 1$ denotes an inverted temperature-density relation, implying that voids are hotter than over-dense regions in the IGM, while $\gamma = 1$ represents an isothermal temperature-density relation. One expects a fiducial value of $\gamma \approx 1.5 - 1.6$ for a post-reionization IGM (Hui & Gnedin, 1997; McQuinn & Upton Sanderbeck, 2016).

35

6.3 Ionization Equilibrium of the Intergalactic Gas

The intergalactic gas is in ionization equilibrium, when the number of photoionizations balances the number of recombinations:

$$n_e \, n_p \, \alpha_{\mathrm{HII}}(T) = n_{\mathrm{HI}} \Gamma_{\mathrm{HI}}, \tag{6.2}$$

where $\alpha_{\mathrm{HII}}(T)$ is the temperature dependent recombination rate, and Γ_{HI} represents the total ionization rate, which is a combination of the ionization field from the close-by quasar Γ_{QSO}, the ionizing background field Γ_{UVB}, and possibly other ionizing sources. The parameters n_e, n_p, and n_{HI} denote the number density of free electrons, protons, and neutral hydrogen in the IGM, respectively.

Following Khrykin et al. (2016)[1], the timescale on which ionization equilibration is reached, can be inferred from the time evolution of the neutral gas fraction x_{HI} of the IGM at a fixed gas temperature, which evolves as

$$\frac{\mathrm{d}x_{\mathrm{HI}}}{\mathrm{d}t} = -\Gamma_{\mathrm{HI}} x_{\mathrm{HI}} + \alpha_{\mathrm{HII}} n_e (1 - x_{\mathrm{HI}}). \tag{6.3}$$

Assuming $\Gamma_{\mathrm{HI}} \sim$ const., Eqn. (6.3) is solved by

$$x_{\mathrm{HI}}(t_{\mathrm{Q}}) = x_{\mathrm{HI},0} + (x_{\mathrm{HI},0} - x_{\mathrm{HI,eq}}) \exp^{-t_{\mathrm{Q}}/t_{\mathrm{eq}}}, \tag{6.4}$$

where t_{Q} is the quasar lifetime, $x_{\mathrm{HI},0} \approx n_e \alpha_{\mathrm{HII}}/\Gamma_{\mathrm{UVB}}$ is the neutral fraction of the IGM before the quasar turned on, and $x_{\mathrm{HI,eq}} \approx n_e \alpha_{\mathrm{HII}}/(\Gamma_{\mathrm{UVB}} + \Gamma_{\mathrm{QSO}})$ is the new lower neutral fraction that the IGM will reach once ionization equilibrium is attained. The characteristic timescale to reach ionization equilibrium, the equilibration timescale t_{eq}, is

$$t_{\mathrm{eq}} \approx \frac{1}{\Gamma_{\mathrm{UVB}} + \Gamma_{\mathrm{QSO}}} \,, \tag{6.5}$$

which is of the order $t_{\mathrm{eq}} \sim 3 \times 10^4$ yr at a distance corresponding to the size of the proximity zone of a typical quasar with magnitude $M_{1450} \sim -27$ at $z \sim 6$, i.e. $R_p \sim 5$ pMpc. The hydrogen ionization rate at this distance is $\Gamma_{\mathrm{HI}} \sim 10^{-12}\,\mathrm{s}^{-1}$.

6.4 The Opacity of the Intergalactic Medium

The cross section of the Lyα absorption is so large that the IGM is opaque even if its neutral gas fraction is as small as $x_{\mathrm{HI}} \sim 10^{-4}$ (Gunn & Peterson, 1965). We will

---

[1]While Khrykin et al. (2016) report the analogous equation for the He II singly ionized fraction in the context of He II Lyα proximity zones at $z \sim 3$, the exact same set of arguments applies here for neutral hydrogen around $z \sim 6$ quasars.
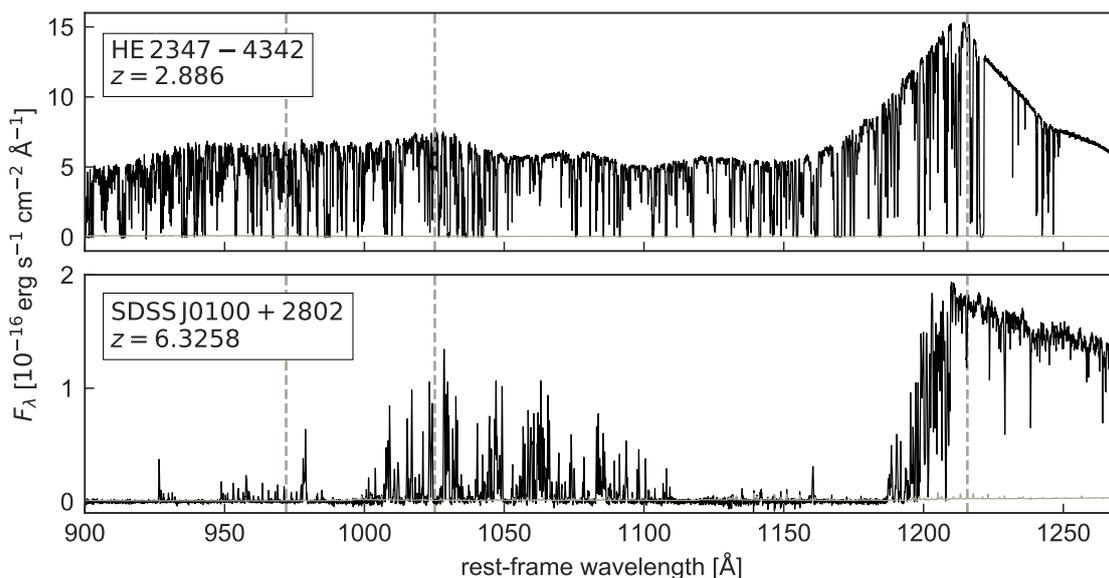
Figure 6.2 Two quasar spectra at medium, i.e. $z \sim 3$ (*top*), and high, i.e. $z \sim 6$ (*bottom*), redshift. The neutral fraction of the IGM increases with redshift while approaching the Epoch of Reionization, resulting in an increase in the optical depth and more absorption in the spectra. The grey dashed vertical lines indicate the Lyα, Lyβ, and $Ly\gamma$ lines at rest-frame wavelengths $\lambda_{\mathrm{Ly}\alpha} = 1215.7$ Å, $\lambda_{\mathrm{Ly}\beta} = 1025.7$ Å, and $\lambda_{\mathrm{Ly}\gamma} = 972.5$ Å, respectively.

define the IGM opacity by means of the effective optical depth, i.e.

$$\tau_{\mathrm{eff}} = -\ln \langle F \rangle, \tag{6.6}$$

where $\langle F \rangle$ is the continuum normalized flux of the quasar or galaxy spectrum, averaged over a spectral region of 50 cMpc h^{-1}.

The evolution of the Lyα forest opacity (see Fig. 6.3), observed in the spectra of $z \gtrsim 6$ quasars, provides the most compelling constraints to date on the end of the epoch of reionization (Fan et al., 2006c; Becker et al., 2015b; Bosman et al., 2018; Eilers et al., 2018a). The detection of transmitted flux spikes in the high redshift quasar spectra and the absence of large, completely opaque, spectral regions, so-called Gunn-Peterson (GP) troughs (Gunn & Peterson, 1965), below $z \lesssim 5$ indicate that the reionization epoch must be completed by that time. Both the steep rise in the observed opacity around $z \gtrsim 5.5$ as well as the increased scatter of the measurements suggest a qualitative change in the ionization state of the IGM (Becker et al., 2015b), provoked by a decrease in the ionizing UVB radiation (Calverley et al., 2011; Wyithe & Bolton, 2011).

The inferred rapid decline of the UVB radiation has been interpreted as an indication for the end stages of reionization (e.g. Fan et al., 2006c; Bolton & Haehnelt, 2007c; Calverley et al., 2011). Fan et al. (2006c) argue further that the increased
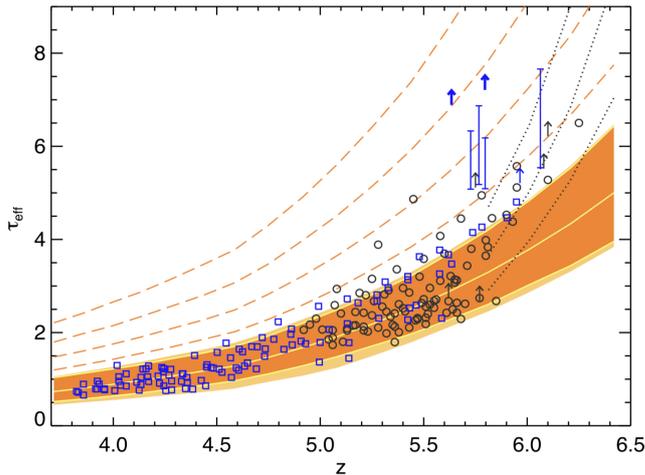
Figure 6.3 Estimated redshift evolution of the Lyα forest opacity by Becker et al. (2015b). The blue data points show their measurements and the dark shaded orange region represents the predicted evolution from their simulations, in which fluctuations in the underlying density field alone account for the scatter in τ_{eff}.

scatter in the opacity measurements around $z \gtrsim 5.5$ could be explained by strong variations in the UVB radiation field as expected in patchy and inhomogeneous reionization scenarios.

However, Lidz et al. (2006) argued that large scale density fluctuations alone could explain the significant variations between sightlines. They calculated the scatter arising solely from density fluctuations while assuming a uniform UVB, which gives results comparable to the observations by Fan et al. (2006c). If this was correct, the evidence for patchy reionization based on the observations of the averaged opacity would be significantly weakened.

Measurements of the Lyα forest opacity along additional quasar sightlines by Becker et al. (2015b) finally showed that the observed scatter in the optical depth measurements significantly exceeds not only fluctuations expected from the density field alone, but also fluctuating UVB models with a spatially-uniform mean free path of ionizing photons (Fig. 6.3). They posited that, if a fluctuating UVB was in fact the source of the large scatter in optical depth, the mean free path must be spatially variable, supporting the interpretation of probing the end stages of an inhomogeneous hydrogen reionization period. Subsequently, Davies & Furlanetto (2016) modeled the UVB with a spatially-varying mean free path and found that the additional fluctuations were sufficient to explain the extra scatter in the optical depth measurements (see also Chardin et al., 2017).

An alternative explanation for the observed scatter in the mean opacity was presented by D'Aloisio et al. (2015), who showed that residual spatial fluctuations in the temperature field could result in extended opacity variations. The spatially varying temperature field is a natural consequence of an extended and inhomogeneous reionization process, wherein regions that reionized early have had time to cool, while regions that reionized late are still hot. The amplitude of the resulting

opacity variations then depends directly on the timing and duration of the reionization process.

In an attempt to disentangle the two competing models of the fluctuating UVB or the fluctuating temperature field, Becker et al. (2018) observed the galaxy population in the vicinity of the longest and most opaque GP trough at $z < 6$ known to date. While in one scenario a weak UVB field due to a deficit of galaxies would produce the observed GP trough, the residual temperature fluctuations in the competing scenario would predict an overdensity of galaxies in the vicinity of the GP trough (Davies et al., 2018a). They observe an underdensity in the population of Lyα-emitting galaxies surrounding GP trough, and conclude that the observed scatter in the IGM opacity is primarily driven by large-scale UVB fluctuations.

In Chapter 9 we will infer the redshift evolution of the IGM opacity from a new and homogeneous data set of high redshift quasar spectra and derive new constraints about the timing and morphology of the Epoch of Reionization.

6.5 Quasar Proximity Zones

Luminous quasars exhibit a region of enhanced flux transmission immediately blueward of the Lyα emission line, which is the so-called quasar proximity zone, before the onset of near complete Gunn-Peterson absorption. This enhanced transmission is caused by the radiation from the quasar itself, which ionizes the surrounding IGM (see e.g. Madau & Rees, 2000; Cen & Haiman, 2000; Haiman & Cen, 2001; Wyithe et al., 2005; Bolton & Haehnelt, 2007b,a; Lidz et al., 2007; Bolton et al., 2011; Keating et al., 2015; Eilers et al., 2017a).

Haiman & Cen (2001) showed that in a very simplistic picture of reionization in which isolated ionized H II regions, known as *Strömgren spheres* (Strömgren, 1939), expand into a neutral uniform ambient IGM, the (proper) size R_{ion} of the ionized region around the quasar depends on the neutral hydrogen fraction x_{HI}, the rate at which ionizing photons are emitted $\dot{N}_\gamma$, and the quasar lifetime t_Q:

$$R_{\text{ion}} \approx \left(\frac{3\dot{N}_\gamma t_Q}{4\pi n_H x_{\text{HI}}} \right)^{1/3}, \tag{6.7}$$

where n_H is the number density of hydrogen. In Eqn. (6.7) recombinations of the ionized gas inside the H II region are neglected, which should be unimportant on the timescale that the quasar turned on, because the recombination timescale of hydrogen is comparable to the Hubble time in the low density IGM. However, there are a number of other effects that are not taken into consideration in this relation,

such as overlapping ionized H II regions, large-scale structure effects, and pre-ionization by local galaxies or the clumpiness of the IGM (see also Fan et al., 2006c; Bolton & Haehnelt, 2007b,a; Lidz et al., 2007; Maselli et al., 2007, 2009; Khrykin et al., 2016).

While Eqn. (6.7) provides a reasonable description of the expansion rate of an H II ionization front embedded in a neutral IGM, it does not predict the distribution of residual neutral hydrogen within the ionized bubble. However, the *observed* size of the proximity zone R_p will depend sensitively on the fraction and distribution of the residual neutral hydrogen, because this can cause saturated absorption well before the ionization front is reached (Bolton & Haehnelt, 2007b; Maselli et al., 2009; Eilers et al., 2017a). Thus the extent of the ionized H II bubble around the quasar can be significantly larger than the observed proximity zone, which is defined to end where the transmitted flux drops below the 10% level (see e.g. Fan et al., 2006c; Carilli et al., 2010) and hence does not necessarily probe the location of the ionization front.

Furthermore Bolton & Haehnelt (2007b) showed that a quasar embedded in a highly ionized IGM can produce a proximity zone that appears qualitatively similar to that of one in a neutral ambient IGM. They show that the observed sizes of proximity zones within an already ionized surrounding IGM are independent of the neutral gas fraction of the ambient IGM and as such, may not provide insights into the evolution of the neutral hydrogen fraction during the Epoch of Reionization at all, an assumption we will confirm in Chapter 11.

However, previous observational studies of the sizes of quasar proximity zones found evidence for a steep decrease in proximity zone size with increasing redshift within the redshift range of $5.7 \le z \le 6.4$ (Fan et al., 2006c; Willott et al., 2007, 2010; Carilli et al., 2010). Although there is large scatter in the observations, this has been interpreted as a strong evolution of the neutral gas fraction of the IGM assuming that the observed proximity zones trace the extent of the ionized H II region presented in Eqn. (6.7), i.e. $R_p \approx R_{\mathrm{ion}}$. However, the measurements of the proximity zone sizes contain a number of uncertainties, most importantly the large uncertainties in the systemic redshifts of the quasars and the resulting ambiguity in the beginning of the proximity zone.

The extent of the H II region around the quasar additionally depends on the lifetime of the quasar, as shown in Eqn. (6.7). Even if the IGM is highly ionized prior to the quasar turning on, the gas in the ionized region responds on a finite timescale (see § 6.3) and therefore also the observed proximity zone sizes depend on the quasar lifetime. Thus, assumptions about the quasar lifetimes are required

to interpret quasar proximity zones, but to date this value remains uncertain by several orders of magnitude (see § 5.3).

In Chapter 11 we will measure the sizes of proximity zones for a new data set of quasars at high redshift and show how they can provide new constraints on the lifetime of these quasars, rather than on the neutral gas fraction of the surrounding IGM.

Structure of this Thesis

In this thesis we study various aspects of the universe by modeling spectroscopic data. In part II we will look in the distant past of our universe and study quasars at a time when the universe was only a few hundred thousand years old. From the analysis of the spectra of these high redshift quasars we will derive new constraints on the Epoch of Reionization, as well as on the growth of SMBHs. Afterwards, in part III, we turn to the local universe and study the dynamics of the Milky Way traced by luminous red giant stars. Spectroscopic observations from various telescopes around the Earth will be the groundwork of all studies that we present.

Chapter 8 introduces a new data set of quasar spectra at high redshift, i.e. $5.77 \lesssim z_{\mathrm{em}} \lesssim 6.54$, observed with the Echellette Spectrograph and Imager (ESI) on the Keck telescopes in Hawaii, which we will use predominantly to analyze the distant universe. This chapter includes work from Eilers et al. (2017a) and Eilers et al. (2018a), both of which are published in the *Astrophysical Journal*.

Chapter 9 uses this data set and shows new constraints on the timing and morphology of the Epoch of Reionization based on measurements of the opacity of the IGM within the Lyα forest in these quasar spectra. The results of this study have been published in Eilers et al. (2018a) in the *Astrophysical Journal*.

Chapter 10 builds upon the results of the previous chapter and provides further constraints on the reionization epoch by measuring the opacity within the Lyβ forest, which is more sensitive to low neutral gas fractions than Lyα due to the higher oscillator strength of the Lyβ transition. This chapter is based on work that will be published in an upcoming paper (Eilers et al., to be submitted to the *Astrophysical Journal*).

In **Chapter 11** we measure the sizes of proximity zones around quasars and introduce a novel method to constrain lifetimes of individual quasars. We identify

three very young quasars in our data set with lifetimes of only $t_Q \sim 10^4 - 10^5$ yr, which pose a significant challenge to all current black hole formation theories. This work has been published in Eilers et al. (2017a) in the *Astrophysical Journal*.

Based on these results **Chapter 12** presents the first spectroscopic study of the youngest quasar known to date, making use of new optical and NIR spectroscopic observations, with the aim to investigate whether there are any spectral signatures that might distinguish the young objects from the quasar population at large. This study has been published in the *Astrophysical Journal* in Eilers et al. (2018b).

In **Chapter 13** we move towards the local universe and summarize a new data-driven method to infer precise parallaxes to luminous red giant stars in the Milky Way, by means of their multi-band photometric and spectroscopic features. This chapter is based on Hogg, Eilers, & Rix (2018), and has been submitted to the *Astrophysical Journal*.

In **Chapter 14** we apply these spectrophotometric distances and measure the most precise circular velocity curve of the Milky Way to date. This work has been published in the *Astrophysical Journal* in Eilers et al. (2019).

We summarize the main results of this thesis and present several projects and ideas for future work in **Chapter 15** and **Chapter 16**.

Part II

Quasars in the Early Universe

A New Data Set of High Redshift Quasar Spectra

Several projects in this thesis are based on a new data set of high redshift quasar spectra, which we will introduce in this chapter. We will describe the observations and the data reduction procedure in § 8.1 and § 8.2, respectively. The properties of this quasar sample are described in § 8.3, before presenting the public data release of the spectra in § 8.4. At the end of this chapter, in § 8.5, we describe our method to normalize the quasar spectra by their continuum emission.

8.1   Observations

Our complete data set consists of 34 quasar spectra at redshifts $5.77 \leq z_{\mathrm{em}} \leq 6.54$, observed at optical wavelengths (4000 Å - 10000 Å) with ESI (Sheinis et al., 2002) at the Keck II Telescope in the years 2001 to 2016. The data was collected from the Keck Observatory Archive (KOA)[1] and complemented with our own observations of four objects (PSO J036 + 03, PSO J060 + 25, SDSS J0100 + 2802, and SDSS J1137 + 3549), that we observed in January of 2016. All observations were obtained using slit widths ranging from $0.75" - 1.0"$, resulting in a resolution of $R \approx 4000 - 5400$. The total exposure times vary from 0.3 h $\lesssim t_{\mathrm{exp}} \lesssim 25$ h resulting in median signal-to-noise ratios in the quasar continuum, at rest-frame wavelengths of 1250Å-1280Å between $2 \lesssim \mathrm{S/N} \lesssim 120$ per pixel. The details of the individual observation runs are shown in the Appendix in Table A.1.

---

[1]https://koa.ipac.caltech.edu/cgi-bin/KOA/nph-KOAlogin

8.2 Data Reduction

We reduce all spectra uniformly using the ESIRedux pipeline[2] developed as part of the XIDL[3] suite of astronomical routines in the Interactive Data Language (IDL). This pipeline employs standard data reduction techniques which can be summarized as follows: Images are overscan subtracted, flat fielded using a normalized flat field image, and then wavelength calibrated using a wavelength image constructed from afternoon arc lamp calibration images. After identifying the objects in the science frames, we subtracted the background using B-spline fits (Kelson, 2003; Bernstein et al., 2015) to the object free regions of the slit. The profiles of the science objects were also fit with B-splines, and an optimal extraction was performed on the sky-subtracted frames. One-dimensional spectra of overlapping Echelle orders are combined to produce a final spectrum for each exposure, and individual exposures are co-added into our final one-dimensional spectra. See Bochanski et al. (2009) for a more detailed description of the algorithms used.

We further optimized the XIDL ESI pipeline to improve the data reduction for quasars at high redshift by differentiating two images (ideally taken during the same observing run) with similar exposure times, analogous to the standard difference imaging techniques performed for NIR observations, in order to improve the sky-subtraction especially in the reddest Echelle orders, which are affected by fringing. This procedure requires dithered exposures for which the trace of the science object lands at different spatial locations on the slit. Dithered exposures have the additional advantage that different parts of the fringing pattern are being sampled, and hence a combination of different exposures further reduces fringing issues in the data. However, since not every observer dithered their object along the slit it was not possible for us to apply this procedure to approximately 10% of the exposures that we took from the archive.

We co-added exposures from different observing runs taken by various principal investigators (PIs) to maximize the signal-to-noise ratio. To combine the data from different observing runs, we weighted each one-dimensional spectrum by its squared signal-to-noise ratio (S/N^2), determined in the quasar continuum region of each spectrum, i.e. at wavelengths longer than the Lyα emission line. This ensures that spectral regions with low or no transmitted flux, which are common in high redshift quasar spectra, obtain the same weight as regions with more transmitted flux.

---

[2]http://www2.keck.hawaii.edu/inst/esi/ESIRedux/
[3]http://www.ucolick.org/~xavier/IDL/
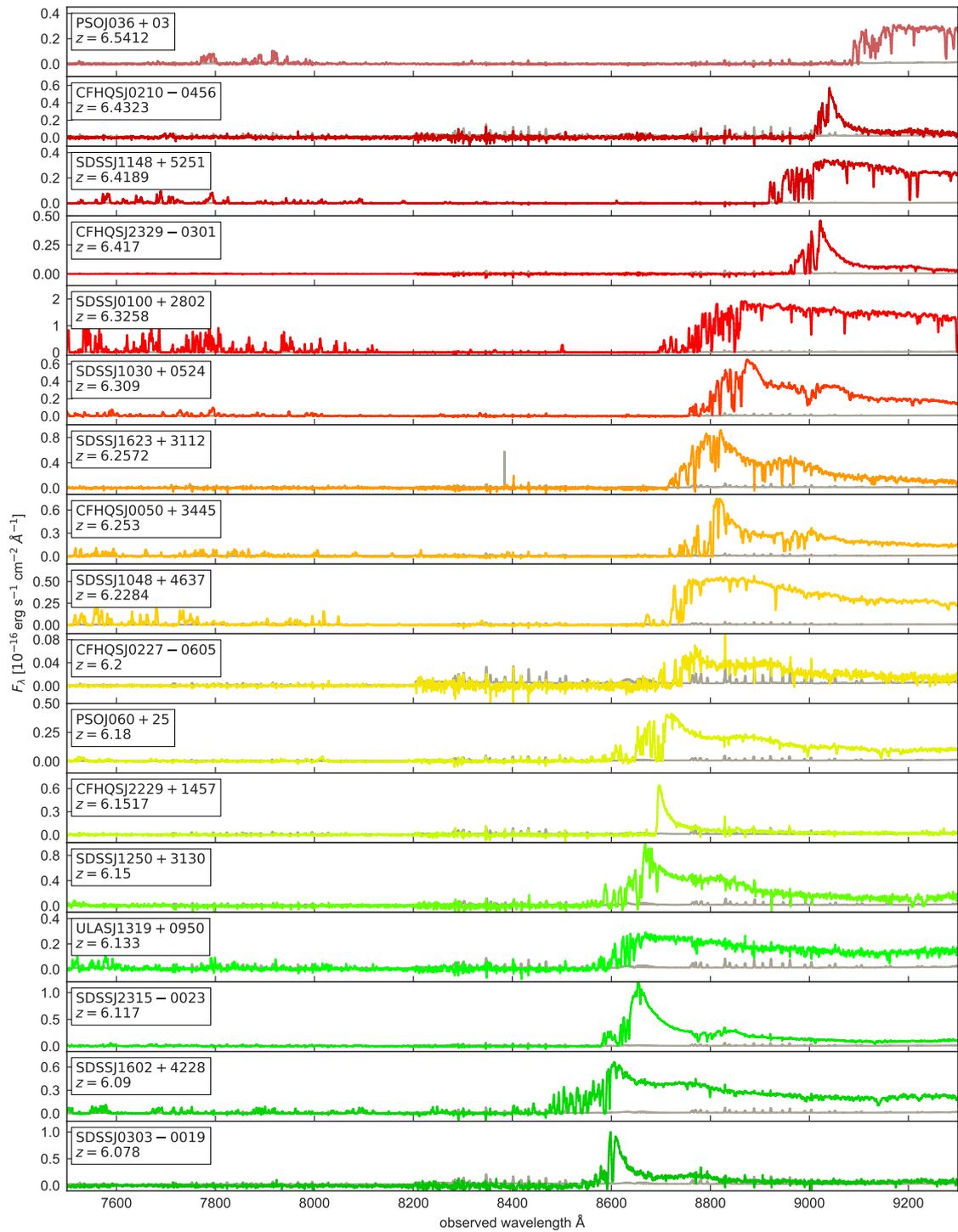
Figure 8.1 Spectra of all quasars in our data sample sorted by redshift.

All final reduced quasar spectra are shown in Fig. 8.1 and 8.2, sorted by their emission redshift.
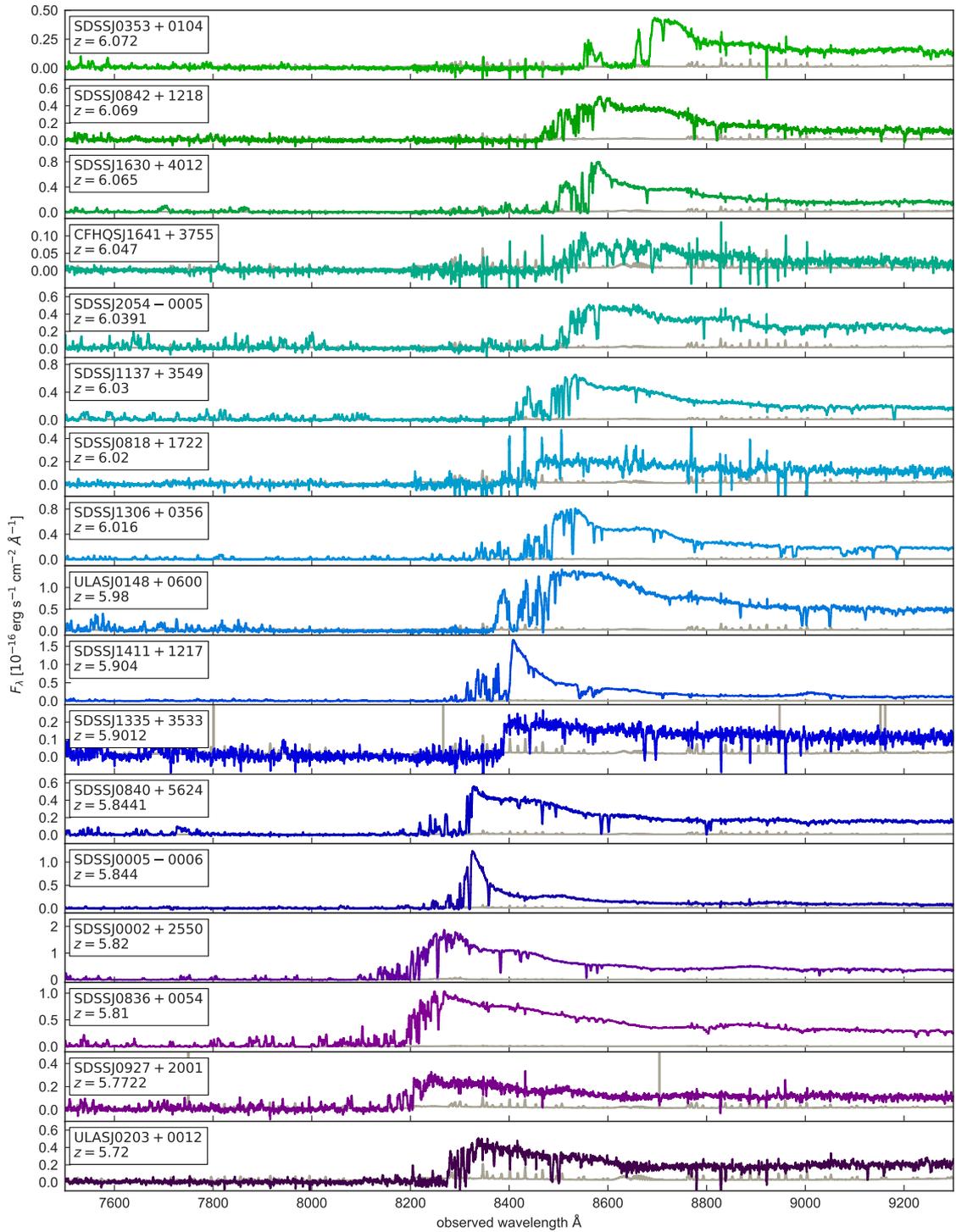
Figure 8.2 Continuation of Fig. 8.1.

8.3   Quasar Properties

The properties of all quasars in our data sample, such as the estimated emission redshift z_{em} and its source, the absolute magnitude M_{1450} at $\lambda_{\text{rest}} = 1450$ Å in the

rest frame, and the S/N ratio of their spectra, are summarized in Table 8.1.

We take the absolute magnitudes M_{1450} defined at $\lambda_{\mathrm{rest}} = 1450$ Å in the rest-frame from Bañados et al. (2016), who determined M_{1450} for all sources in a consistent way. They assume a power law continuum slope $\alpha_\nu = -0.3$ and require the extrapolated y_{P1}- or J-band magnitudes (at $\lambda_{\mathrm{eff}} = 9627.7$ Å or $\lambda_{\mathrm{eff}} = 12444.0$ Å, respectively) to be consistent with their measurements, since most optical quasar spectra at these high redshifts have limited wavelength coverage at $\lambda_{\mathrm{rest}} = 1450$Å.

8.3.1 Redshifts

Determining precise redshifts for quasars is very challenging due to the broad widths of emission lines, Gunn-Peterson absorption, and offsets between different ionization lines (Gaskell, 1982; Tytler & Fan, 1992; Vanden Berk et al., 2001; Richards et al., 2002; Shen et al., 2016). Most quasars show strong internal motions and winds, which displace many of the emission lines, such as the Lyα line or far-UV lines, far from the systemic redshift of the host galaxy. Thus the most precise determination of the location of the quasar are low ionization lines such as Mg II lines or, even better, emission lines from the molecular gas reservoir of the host galaxy itself, such as CO or [C II] lines.

We account for uncertainties in the systemic redshift of each quasar depending on the measured emission lines to determine their redshifts. For redshifts determined from the detection of a [C II] line at $158\,\mu$m or a CO line we assign a redshift error of $\Delta v = 100\,\mathrm{km\,s}^{-1}$. For quasars with a redshift measurement from a Mg II line we assume a redshift error of $\Delta v = 270\,\mathrm{km\,s}^{-1}$, in order to account for the dispersion between the redshift of the Mg II line and the redshift of the host galaxy (e.g. Richards et al., 2002; Hennawi et al., 2006; Shen & Ménard, 2012; Venemans et al., 2016; Shen et al., 2016), and for the remaining quasars for which the redshift was determined by the Lyα line or far-UV lines, which usually suffer from large velocity offsets, we assume a redshift error of $\Delta v = 1000$ km/s. Note that these uncertainties are chosen to be very conservative (Shen et al., 2016), because the redshift measurements are taken from the literature, measured by various authors possibly using different data and methods.

8.4 Data Release

Our new data set including the final co-added spectra and their noise vectors together with the estimated quasar continua (see § 8.5), as well as additional

Table 8.1 Overview of the properties of the quasar sample.

object	z_{em}	Δv [km s^{-1}]	redshift line	Ref.[a]	M_{1450}	S/N[b]
SDSS J0002+2550	5.82	1000	Lyα -O I	1	−27.31	62
SDSS J0005-0006	5.844	270	Mg II	10	−25.73	15
CFHQS J0050+3445	6.253	270	Mg II	9	−26.70	18
SDSS J0100+2802	6.3258	100	[C II]	17	−29.14	41
ULAS J0148+0600	5.98	270	Mg II	14	−27.39	30
CFHQS J0210-0456	6.4323	100	[C II]	12	−24.53	2
PSO J036+03	6.5412	100	[C II]	13	−27.33	10
CFHQS J0227-0605	6.20	1000	Lyα	6	−25.28	3
SDSS J0303-0019	6.078	270	Mg II	7	−25.56	2
PSO J060+24	6.18	1000	Lyα	16	−26.95	13
SDSS J0818+1723	6.02	1000	Lyα	2	−27.52	8
SDSS J0836+0054	5.810	270	Mg II	4	−27.75	120
SDSS J0840+5624	5.8441	100	CO	8	−27.24	28
SDSS J0842+1218	6.069	270	Mg II	10	−26.91	10
SDSS J0927+2001	5.7722	100	CO	8	−26.76	7
SDSS J1030+0524	6.309	270	Mg II	3	−26.99	26
SDSS J1137+3549	6.03	1000	Lyα -O I -Si IV	2	−27.36	24
SDSS J1148+5251	6.4189	100	[C II]	15	−27.62	34
SDSS J1250+3130	6.15	1000	Lyα -O I -Si IV	2	−26.53	8
SDSS J1306+0359	6.016	270	Mg II	4	−26.81	41
ULAS J1319+0950	6.1330	100	[C II]	11	−27.05	10
SDSS J1335+3533	5.9012	100	CO	8	−26.67	7
SDSS J1411+1217	5.904	270	Mg II	4	−26.69	32
SDSS J1602+4228	6.09	1000	Lyα -O I	1	−26.94	21
SDSS J1623+3112	6.2572	100	[C II]	18	−26.55	9
SDSS J1630+4012	6.065	270	Mg II	7	−26.19	15
CFHQS J1641+3755	6.047	270	Mg II	9	−25.67	4
SDSS J2054-0005	6.0391	100	[C II]	11	−26.21	17
CFHQS J2229+1457	6.1517	100	[C II]	15	−24.78	2
SDSS J2315-0023	6.117	1000	Lyα	5	−25.66	15
CFHQS J2329-0301	6.417	270	Mg II	9	−25.25	2

Notes. The columns show the object name, the redshift of the quasar and the redshift uncertainty noted as a velocity offset, the lines measured to determine the redshift and the reference therefore, the quasar's magnitude M_{1450} and the S/N of the spectrum.

[a] Reference for redshift. 1: Fan et al. (2004), 2: Fan et al. (2006b), 3: Jiang et al. (2007), 4: Kurk et al. (2007), 5: Jiang et al. (2008), 6: Willott et al. (2009), 7: Carilli et al. (2010), 8: Wang et al. (2010), 9: Willott et al. (2010), 10: De Rosa et al. (2011), 11: Wang et al. (2013), 12: Willott et al. (2013), 13: Bañados et al. (2015), 14: Becker et al. (2015b), 15: Willott et al. (2015), 16: Bañados et al. (2016), 17: Wang et al. (2016), 18: private communication with R. Wang.

[b] Median S/N per pixel; estimated between 1250 Å$\leq \lambda_{\rm rest} \leq$ 1280 Å.

meta data on the quasars, are available in the third data release of the `igmspec` database[4] (Prochaska, 2017), and via the `zenodo`[5] platform. A catalog for the data release comprising the main properties of the data set is shown in Tab. A.2 in Appendix A.

8.5 Quasar Continuum Normalization

All quasar spectra were normalized to unity at $\lambda_{\rm rest} = 1280$ Å in a spectral region that is free of emission lines. The quasar continuum was then estimated with principal component spectra (PCS) from a Principal Component Analysis (PCA) of lower redshift quasar spectra (Pâris et al., 2011). The idea of the PCA is to represent each continuum spectrum $|q_\lambda\rangle$ at wavelength λ by a reconstructed spectrum comprising a mean spectrum $|\mu_\lambda\rangle$ and a sum of m weighted PCS $|\xi_\lambda\rangle$, i.e.

$$|q_\lambda\rangle \approx |\mu_\lambda\rangle + \sum_{i=1}^{m} \alpha_i |\xi_{i,\lambda}\rangle, \tag{8.1}$$

where the index i refers to the ith PCS and α_i denotes its weight.

Since the quasars in our data sample are all at $z_{\rm em} \sim 6$, they experience substantial absorption due to intervening neutral hydrogen within the IGM bluewards of the Lyα emission line. Thus the continuum estimate was performed solely on the unabsorbed continuum spectrum redwards of the Lyα emission line with a set of PCS from Pâris et al. (2011) covering the wavelengths 1215.67 Å$\leq \lambda_{\rm rest} \leq$ 2000 Å. We take the model that minimizes χ^2 using the noise vector from the spectra as the best estimate.

In order to obtain coefficients $\boldsymbol{\alpha}$ for a set of PCS that cover the *entire* spectral region between 1020 Å $\leq \lambda_{\rm rest} \leq$ 2000 Å, we use a projection matrix $\boldsymbol{P}$ to project the estimated coefficients for the PCS *redwards* of Lyα onto this new set of coefficients

---

[4] http://specdb.readthedocs.io/en/latest/igmspec.html
[5] DOI: 10.5281/zenodo.1467821
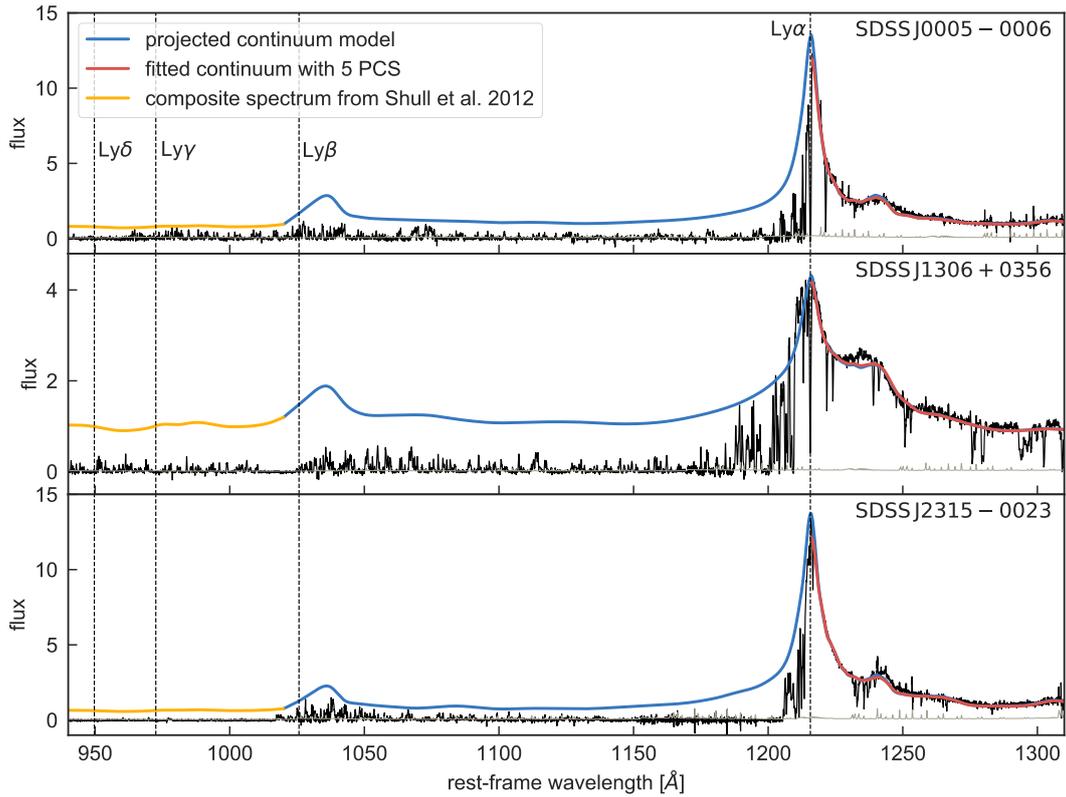
Figure 8.3 Examples of three quasar spectra in our data sample and their continuum models. The continua are first fitted with five PCS from Pâris et al. (2011) to the unabsorbed quasar continuum at wavelength $\lambda_{\mathrm{rest}} \geq 1215.67$ Å (red part) and afterwards projected onto the blue side of the quasar spectrum (blue part). The continuum model is then augmented to lower wavelengths by appending the composite spectrum from Shull et al. (2012) at $\lambda_{\mathrm{rest}} \leq 1020$ Å (yellow part). The vertical dashed lines mark the location of the Lyα, Lyβ, Lyγ, Lyδ emission line.

for the entire spectrum. The projection matrix $\boldsymbol{P}$ has been computed by Pâris et al. (2011) using the set of PCS for both the red wavelength side only and the whole spectral region covering wavelengths bluewards and redwards of Lyα. Hence

$$\boldsymbol{\alpha} = \boldsymbol{P} \cdot \boldsymbol{\alpha}_{\mathrm{red}}. \tag{8.2}$$

This new set of coefficients together with Eqn. (8.1) provides a continuum model for each quasar covering all wavelengths 1020 Å $\leq \lambda_{\mathrm{rest}} \leq 2000$ Å. Pâris et al. (2011) estimate that the median uncertainty on the transmitted flux in the Lyα forest to be $|\Delta F_{\mathrm{forest}}| \approx 5\%$. However, since we do not take all PCA components into account and do not have the full spectral coverage up to $\lambda_{\mathrm{rest}} = 2000$ Å to estimate the continua, the uncertainty on the continua in our quasar spectra is most likely higher, i.e. $|\Delta F_{\mathrm{forest}}| \approx 10 - 20\%$.

For an estimate of the continua at lower wavelength we take the composite

quasar spectrum provided by Shull et al. (2012), constructed from 22 low redshift quasars observed with the COS on the Hubble Space Telescope (HST) that extends from 550 Å to 1750 Å in the rest frame, and re-scale the composite spectrum to match the PCA constructed continuum model at $\lambda_{\mathrm{rest}} = 1020$ Å. We augment the continuum model by simply appending the composite spectrum at wavelengths $\lambda_{\mathrm{rest}} < 1020$ Å. A few example quasar spectra from our data set and their continuum model are shown in Fig. 8.3, all remaining quasar spectra that we analyzed and their estimated continua are shown in Fig. A.1 in the appendix.

The Timing and Morphology of the Epoch of Reionization

In this chapter, we apply our new data set of high redshift quasar spectra that we introduced in Chapter 8 to set constraints on the timing and morphology of the Epoch of Reionization by measuring the evolution of the Lyα opacity of the IGM. We will first describe how we measure the opacity (§ 9.1) and present our measurements in § 9.2. We then introduce a suite of hydrodynamical simulations to compare our results in § 9.3, and discuss their implications on the Epoch of Reionization in § 9.4.

9.1 Measuring the Optical Depth of the Intergalactic Medium

From our data sample of 34 quasar sightlines we exclude the three broad absorption line (BAL) quasars (ULAS J0203+0012, SDSS J0353+0104, and SDSS J1048+4637) to avoid potential contamination by broad non-IGM absorption in the Lyα forest. We further exclude quasars with only very low S/N data, i.e. S/N < 7, whose spectra are easily subject to systematic errors. Our final sample to analyze the IGM optical depth consists of 23 quasar spectra.

As described in § 6.4 we estimate the mean opacity of the IGM by means of the effective optical depth, which is defined in Eqn. (6.6) as the negative natural logarithm of the continuum normalized flux, averaged within discrete spectral bins along the line of sight of each quasar. We chose fixed comoving bins of size 50 comoving Mpc h^{-1} (cMpc h^{-1}) (Becker et al., 2015b), which contains a similar path
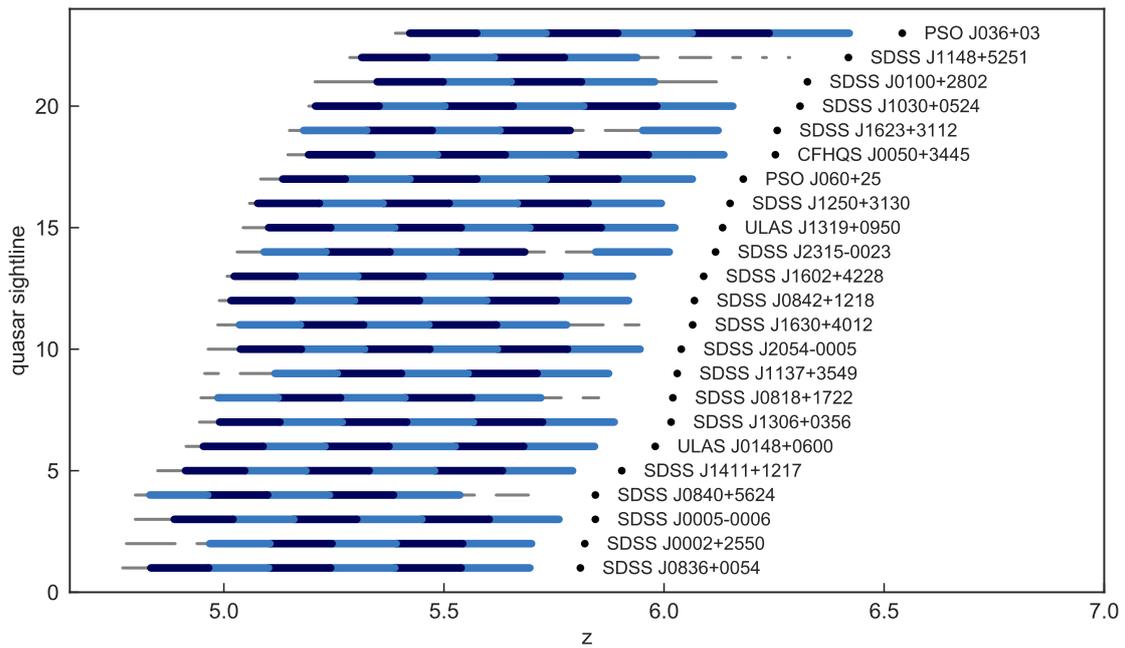
Figure 9.1 Redshift coverage of each quasar spectrum used for the opacity measurements of the IGM. The black dots show the emission redshift of each quasar. The different blue colored regions show the 50 cMpch^{-1} bins within the Lyα forest. The remaining unused coverage of each spectrum is shown in grey. Masked regions are not shown.

length as the bins of size $\Delta z = 0.15$ at $z \sim 5 - 6$ previously applied by Fan et al. (2006c).

In order to avoid biases in the measurement of the opacity of the IGM, we mask the spectral region around each quasar that is strongly influenced and ionized by the quasar's own radiation. We use the measurements for the proximity zones R_p (which will be presented in Tab. 11.1 in Chapter 11) as an estimate the influenced region. However, the proximity zone is defined such that it does not completely reach out to the ionization front expanding from the quasar. Thus the influence of the quasar's radiation is expected to be still present outside of its measured proximity zone (Eilers et al., 2017a), since the radiation of the quasar still dominates the UVB radiation, i.e. $\Gamma_{\text{QSO}} \gg \Gamma_{\text{UVB}}$ at R_{p}. Hence, we mask an additional 2.5 pMpc around each quasar, i.e. the masked region measures $R_{\text{p}} + 2.5$ pMpc, in order to eliminate all enhanced transmission due to the quasar's radiation.

Thus we choose the maximum wavelength that we consider for opacity measurements to lie just blueward of this masked region. The minimum wavelength we consider for measurements within the Lyα forest is 1030 Å in the rest frame, in order to account for possible redshift uncertainties.

Another possible contamination in the measurement of optical depths are in-

tervening absorbers along the line of sight, such as Damped Lyα Systems (DLAs) or other low-ion metal absorbers which are likely associated with relatively high H I column density ($N_{\mathrm{HI}} \gtrsim 10^{19}$ cm^{-2}). We mask the regions in the quasar spectra around these absorbers, as they reflect an absorption signature that is not typically resolved in IGM simulations. To this end, we searched for low-ion metal absorption lines, such as e.g. Al II, Fe II, and O I, associated with absorbers in the continuum spectra redwards of the Lyα emission line that are located at the same redshift as a spectral region showing complete absorption in the forest of the spectrum. Additionally, we searched through the literature for DLAs and low-ion metal absorbers along the quasar sight lines in our sample. For each absorber, we conservatively masked the spectral region around the absorption system within ± 30 Å in the observed wavelength frame in the Lyα forest at the corresponding wavelength. Spectral bins containing absorbers were then excluded from the IGM opacity measurements. Table 9.1 shows a compilation of all identified absorbers along the line of sight to the quasars in our sample. Note that most of the identified absorbers have been already found by other authors, since most quasars in our data set have been previously known.

Additionally, we mask all spuriously high pixels within the Lyα forest of the quasar spectra, by checking for single pixels showing $F > 1$ in the continuum normalized spectra. Because sky-subtraction systematics occasionally result in large negative sky-subtraction residuals, we also mask all negative flux pixels with the 2.5% lowest S/N, to avoid biases due to large uncertainties in pixels that fall onto sky lines and have large negative residuals.

After masking all low-ion metal absorption systems, proximity zones, and spuriously high and negative pixels, the combined usable path length for the opacity measurements is 6350 cMpc$\,h^{-1}$ for the 23 quasar sight lines in our data sample. The spectral regions in which we measure the mean flux and calculate its effective optical depth are shown for each quasar as the dark and light blue colored bars in Fig. 9.1. Masked regions are shown in white. The grey regions show pathlength that are in principle usable but are not used, because the remaining unmasked region would be smaller than our our chosen bin size of 50 cMpc$\,h^{-1}$.

Correcting for Offsets in the Zero-Level of the Spectra

The noise for pixels with no intrinsic flux should be symmetrically distributed around zero, since pixels with zero flux have equal probability to be scattered into positive or negative values. This idea was applied by McGreer et al. (2011), for example, to estimate the number of so called "dark pixels" that have a flux value

Table 9.1 Intervening low-ion absorption systems along the line of sight to the quasars.

object	z_{em}	z_{abs}	Ref.[a]
SDSS J1148+5251	6.4189	6.0115	1
		6.1312	1
		6.1988	1
		6.2575	1
SDSS J2054-0005	6.0391	5.9776	1
SDSS J2315-0023	6.117	5.7529	1
SDSS J1630+4012	6.065	5.8865	1
SDSS J1137+3549	6.03	5.0124	1
SDSS J1623+3112	6.2572	5.8415	1
SDSS J0840+5624	5.8441	5.5940	2
SDSS J0002+2550	5.82	4.914	3
SDSS J0100+2802	6.3258	6.1437	3
SDSS J0818+1722	6.02	5.7911	1
		5.8765	1

Notes. The columns show the name of the object and its emission redshift, the absorption redshift of the intervening low-ion absorber and the reference therefor.

[a] Reference for low-ion absorbers. 1: Becker et al. (2011), 2: Ryan-Weber et al. (2009), 3: this work.

consistent with zero. However, a detailed inspection of the quasar spectra in our data set reveals that the zero-level in some spectral regions can be slightly biased, i.e. we do not observe a symmetric distribution around zero in the noise, possibly caused by sky-subtraction systematics present in a small fraction of exposures. But even tiny offsets in the zero-level can cause large differences in the opacity estimates, especially in regions with very little transmitted flux. Because these are the regions we are particularly interested in, we correct for small offsets in the zero-level of the spectra. These offsets are calculated and applied to each $50\,\mathrm{cMpc}\,h^{-1}$ spectral bin individually in order to avoid correlations between the optical depth measurements.

A detailed description of this correction procedure can be found in Appendix B. We estimate that the systematic error in the mean flux $\langle F^{\mathrm{Ly}\alpha}\rangle$ due to corrections of the zero-level offset is

$$\sigma_{\langle F^{\mathrm{Ly}\alpha}\rangle} \approx 0.0067,$$
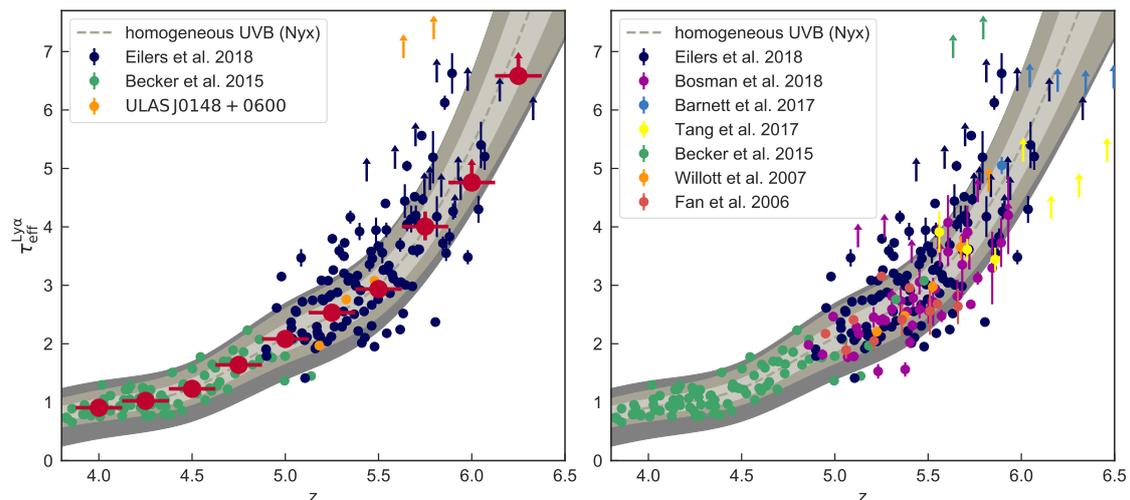
Figure 9.2 Evolution of the Lyα forest effective optical depth. *Left panel:* The dark blue data points show our IGM opacity measurements. The green and orange data points are measurements of the optical depth performed by Becker et al. (2015b), orange indicating the measurements of ULAS J0148 + 0600. This data set we consider the master compilation sample. The large red data points show the mean redshift evolution averaged over bins of $\Delta z = 0.25$, their uncertainties are determined via bootstrapping. The grey underlying region shows the predicted redshift evolution from radiative transfer simulations assuming a uniform UVB model. We have simulation outputs in steps of $\Delta z = 0.5$ and use a cubic spline function to interpolate the shaded regions between the redshifts of the outputs. The light and medium grey shaded regions indicate the 68th and 95th percentile of the scatter expected from density fluctuations in the simulations, whereas the dark grey region shows any additional scatter due to $\sim 20\%$ continuum uncertainties. *Right panel:* Compilation of all opacity measurements found in the literature along quasar sightlines that are not in our data sample and that have been calculated within similar spectral bin sizes.

i.e. less than $< 1\%$, and thus constitutes only a minor correction to the optical depths measurements.

9.2 The Redshift Evolution of the Optical Depth in the Lyα forest

We compute the effective optical depth τ_{eff} from the measurements of the observed mean flux $\langle F^{\mathrm{obs}} \rangle$ in bins of 50 cMpc h^{-1} using Eqn. (6.6). We list all measurements of the mean flux within the Lyα forest for each bin along all 23 quasar sight lines in our data sample in Tab. C.1 in Appendix C. All spectral bins indicating the respective measurements of $\langle F^{\mathrm{obs}} \rangle$ and $\tau_{\mathrm{eff}}^{\mathrm{Ly}\alpha}$ are shown in Fig. C.1, C.2, and C.3. If

the mean flux is detected with less than 2σ significance or if we measure a negative mean flux, we adopt a lower limit on the optical depth at $\tau_{\mathrm{eff}} = -\ln(2\sigma_{\langle F^{\mathrm{obs}}\rangle})$ consistent with previous works. Note that we do not include the systematic uncertainty on the mean flux ($\sim 10 - 20\%$, see § 8.5) arising from the PCA continuum estimate. This uncertainty on the mean flux would lead to an additional uncertainty on $\tau_{\mathrm{eff}}^{\mathrm{Ly}\alpha}$ of $\sim 2 - 5\%$, when most of the flux is absorbed in the quasar spectra (see also Fig. 7 in Becker et al., 2015b). The results of the optical depth measurements within the Lyα forest, plotted as a function of redshift, are shown in Fig. 9.2.

In Fig. 9.2 our new measurements are shown in dark blue in both panels. For the majority of quasar spectra that we use in our analysis the IGM opacity has been analyzed before. However, we have co-added the data from multiple observation runs (see § 8.2 for details), in order to achieve higher signal-to-noise. However, for one object, ULAS J0148 + 0600, the data obtained by Becker et al. (2015b) with VLT/X-Shooter in a 10 h observation, has a higher S/N ratio than our spectrum. This sightline exhibits a particularly deep GP trough (Becker et al., 2015b), and hence the enhanced S/N ratio results in more stringent opacity limits, representing the strongest fluctuations in the IGM opacity at this redshift. In order to model the IGM fluctuations correctly, it is important to include these outliers (Chardin et al., 2015; D'Aloisio et al., 2015; Davies & Furlanetto, 2016). Thus we construct a master compilation of opacity measurements, presented in the left panel of Fig. 9.2, and replace our optical depth measurements within the Lyα forest along just the sightline of ULAS J0148 + 0600 with the more precise measurements obtained by Becker et al. (2015b) (orange data points). This mainly adds the two most stringent limits at $z = 5.634$ and $z = 5.796$ to our analysis, since the better data quality results in higher sensitivity in the GP troughs. The lower redshift $\tau_{\mathrm{eff}}^{\mathrm{Ly}\alpha}$ measurements for this object are consistent with our measurements. This master compilation is shown in the left panel of Fig. 9.2. Additionally, we also show the lower redshift $\tau_{\mathrm{eff}}^{\mathrm{Ly}\alpha}$ measurements from Becker et al. (2015b) (green data points).

We present the average opacity evolution by calculating the mean flux and the bootstrapped error on the mean in bins of $\Delta z = 0.25$ within the Lyα forest from the master compilation set. We then compute the binned opacity values via $\tau_{\mathrm{eff}} = -\ln\langle F\rangle$, where $\langle F\rangle$ is the mean flux computed in these bins. The uncertainties on the opacity values with uncertainties also determined via bootstrapping are shown as the red data points and tabulated in Tab. 9.2. Similar to the individual τ_{eff} measurements, we adopt a limit if the mean flux in the bin is measured with less than 2σ significance (where σ is here the bootstrap errors on the mean flux).

As mentioned before we do not take any systematic errors on the mean flux

measurements into account that could, for instance, result from uncertainties in the continuum estimation. The dark grey regions give an estimate of the additional scatter expected due to continuum uncertainties of $\sim 20\%$, which are negligible at high redshift, where the transmitted flux is low and the scatter is dominated by fluctuations along different sightlines (Becker et al., 2015b; Eilers et al., 2017b).

The right panel of Fig. 9.2 compares our data set to opacity measurements from additional sightlines from the literature that are not in our data sample. The additional data points come from the sightlines of SDSS J0144 − 0125 and SDSS J1436 + 5007 (Fan et al., 2006c), CFHQS J1509 − 1749 (Willott et al., 2007), ULAS J1120 + 0641 (Barnett et al., 2017), PSO J006.1240 + 39.221 (Tang et al., 2017), and J0323 − 4701, J0330 − 4025, J0410 − 4414, J0454 − 4448, J0810 + 5105, J1257 + 6349, J1609 + 3041, J1621 + 5155, J2310 + 1855 (Bosman et al., 2018). In most of these analyses the bins were chosen to be $\Delta z = 0.15$, following Fan et al. (2006c). This bin size covers roughly the same spectral region as the chosen bin size of $50 \text{ cMpc} \, h^{-1}$ in our analysis and the one by Becker et al. (2015b) at $z \sim 6$, but in the redshift interval of $5 \lesssim z \lesssim 7$, the bin size changes quite significantly. Overall the agreement between the various measurements with our new analysis is good, but we chose not to add these measurements to the master compilation, because of the different pathlengths used to construct the measurements, very low S/N data or the variety of different instruments and data reduction pipelines used to obtain the spectra, which enlarges the systematic uncertainties on these measurements (see § 9.2.1).

9.2.1 Comparison to Other Studies

For several quasar sight lines in our data sample the optical depth has been measured previously by Fan et al. (2006c) and Becker et al. (2015b), and more recently by Bosman et al. (2018). However, the quality of the data and the methods to analyze the data differ. Here, we carry out a detailed comparison of our methods and measurements to previous work, and discuss potential systematic uncertainties and resulting discrepancies in the CDFs of the optical depth.

Estimating Systematic Uncertainties

We compare the measurements of the IGM opacity for the 16 quasar sightlines that are both part of our analysis and the data set of Fan et al. (2006c) and are not BAL quasars. The spectra from Fan et al. (2006c) partially overlap with our data set, but six quasars were observed with a different telescope and instrument (MMT − MMT

Table 9.2 Measurements of the average flux and optical depth within the Lyα forest of our master compilation sample.

z_{abs}	$\langle F \rangle$	$\sigma_{\langle F \rangle}$	$\langle \tau_{\mathrm{eff}}^{\mathrm{Ly}\alpha} \rangle$	$\sigma_{\langle \tau_{\mathrm{eff}}^{\mathrm{Ly}\alpha} \rangle}$
4.0	0.4046	0.0151	0.9049	0.0372
4.25	0.3595	0.0112	1.0230	0.0311
4.5	0.2927	0.0190	1.2286	0.0651
4.75	0.1944	0.0150	1.6381	0.0770
5.0	0.1247	0.0132	2.0818	0.1060
5.25	0.0795	0.0078	2.5321	0.0984
5.5	0.0531	0.0058	2.9357	0.1090
5.75	0.0182	0.0045	4.0057	0.2469
6.0	0.0052	0.0043	>4.7595	—
6.25[a]	-0.0025	0.0007	>6.5843	—

Notes. The columns show the mean redshift z_{abs} of the redshift bins of size $\Delta z = 0.25$, the averaged flux $\langle F \rangle$ and its uncertainty $\sigma_{\langle F \rangle}$ determined via bootstrapping, and the mean optical depth $\langle \tau_{\mathrm{eff}}^{\mathrm{Ly}\alpha} \rangle$ in that redshift bin and its error $\sigma_{\langle \tau_{\mathrm{eff}}^{\mathrm{Ly}\alpha} \rangle}$, also determined via boostrapping.
[a] Note that this redshift bin only contains two measurements.

Red Channel, Hobby-Eberly Telescope (HET), Kitt Peak (KP) – KP 4m MARS) and eight quasars have additional Keck/ESI data since their analysis. We have reduced and stacked all of the Keck/ESI observations from the archive and thus the spectra analyzed in this thesis have an improved quality due to their longer exposure time.

Additionally, our methods to analyze the data differ. For instance, Fan et al. (2006c) applied a power-law to the red side of the quasar spectra to estimate the quasar continua, whereas we estimated the quasar continua by a PCA (see § 8.5). In our analysis we mask all spectral regions containing low-ion metal absorption systems, while in previous work it has been argued that those have a negligible influence and can thus be ignored (Fan et al., 2006c; Becker et al., 2015b).

All these differences contribute to the systematic error of the opacity measurements. We attempted to assess these systematic uncertainties by comparing the results from our analysis to Fan et al. (2006c) along the sight lines that are part of both data sets. To this end we measure the mean flux in the same redshift bins as Fan et al. (2006c) with a fixed bin size of $\Delta z = 0.15$, and compare our measurements to Fan et al. (2006c) in Fig. 9.3. We observe a large scatter in the distribution and a systematic offset towards lower mean flux values in our measurements, much larger than the formal measurement uncertainties. The negative offset is strongest
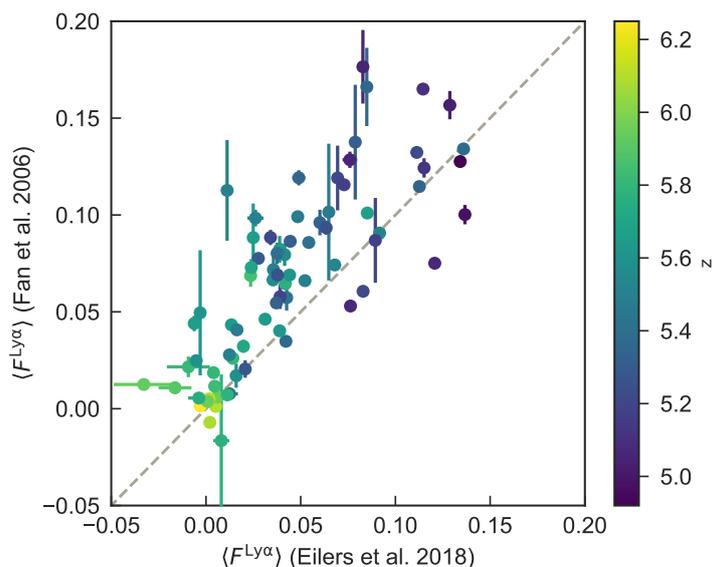
Figure 9.3 Mean flux measurements within the Lyα forest of Fan et al. (2006c) plotted vs. our measurements within the same redshift bins of $\Delta z = 0.15$ for the 16 quasar sightlines that are part of both data sets.

at lower redshifts with higher mean flux values, i.e. lower optical depths.

We estimate the systematic error arising due to different observations, different data reduction pipelines and different analyses by the median of the distribution of measured flux differences $\Delta \langle F \rangle = \langle F_{\mathrm{Eilers\,et\,al.\,2018}} \rangle - \langle F_{\mathrm{Fan\,et\,al.\,2006}} \rangle$. We find a median systematic error of

$$\sigma_{\Delta \langle F \rangle} \approx -0.023,$$

with a large scatter of ≈ 0.026 determined from the mean of the 16th and 84 percentile of the distribution. A detailed investigation of a few of the largest outliers in this distribution suggests that differences in the spectra itself, due to the different instruments with which they were observed and potentially due to differences in the data reduction, cause the largest discrepancies.

A similar analysis of the Lyα optical depths measured from a comparable quasar sample at $z_{\mathrm{em}} > 5.7$ was presented by Bosman et al. (2018). Of the 62 sight lines they analyzed, 22 satisfy our quality criteria, namely that they are non-BAL quasars with a S/N > 7[1]. Out of these, 17 objects overlap with our sample. Although the Bosman et al. (2018) sample is comparable to ours in size and partially overlapping, their methods differ in a variety of important aspects from ours. As in Fan et al. (2006c) different data reduction pipelines have been used to reduce the data, the quasar continuum estimation methods differ, and while both our study and Bosman et al. (2018) mask the proximity zone regions, we adapt the excluded re-

---

[1]Note that Bosman et al. (2018) quote a S/N ratio per $60\,\mathrm{km\,s^{-1}}$ pixel. Our S/N quoted in Tab. 8.1 is calculated per $10\,\mathrm{km\,s^{-1}}$ pixel so for a direct comparison we have to correct our quoted S/N ratios by $\mathrm{S/N}_{60\,\mathrm{km\,s^{-1}}} = \mathrm{S/N}_{10\,\mathrm{km\,s^{-1}}} \cdot \sqrt{6}$, i.e. the threshold we apply for including spectra in our analysis is $\mathrm{S/N}_{60\,\mathrm{km\,s^{-1}}} > 17.1$.
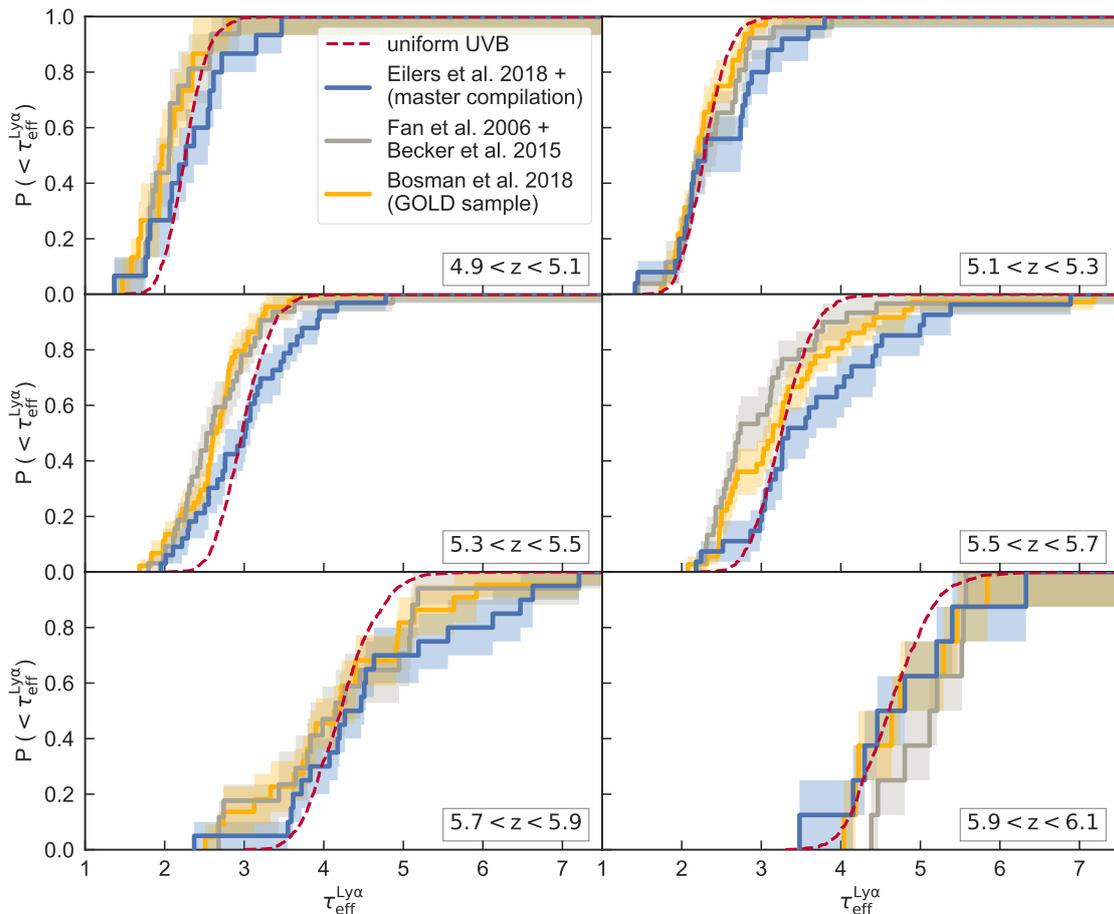
Figure 9.4 CDFs of $\tau_{\mathrm{eff}}^{\mathrm{Ly}\alpha}$ in different redshift bins. The blue curves show the master compilation of our new measurements with ULASJ1048 + 0600 from Becker et al. (2015b), whereas the grey and yellow curves show the CDF from previous measurements (Fan et al., 2006c; Becker et al., 2015b) as well as the new compilation by Bosman et al. (2018), respectively. The shaded regions show the 1σ-uncertainties determined via boostrapping. The red dashed curves show CDFs from hydrodynamical simulations assuming a uniform UVB. Note that the simulations have been re-scaled to match the mean τ_{eff} in the respective bin by applying Eqn. (9.1) and the procedure described in § 9.3.3.

gion dependent on the actual measured proximity zone size R_p (see Chapter 11), whereas their analysis excludes a fixed spectral range until $\lambda_{\mathrm{rest}} = 1178$ Å, which corresponds to $\Delta R_p = 13.3$ pMpc at $z = 6$. Finally, we have masked strong absorbers and account for small zero-level offsets, whereas they do not.

Comparison of the Cumulative Distribution Functions

In Fig. 9.4 we compare the CDF from our measurements shown in blue to the CDF from previous studies by Fan et al. (2006c) and Becker et al. (2015b), which are

shown as the grey curves, and by Bosman et al. (2018) shown in yellow, in different redshift bins centered around $5.0 \leq z \leq 6.0$. We show the so-called GOLD sample from Bosman et al. (2018) including 33 quasar spectra for which they applied a data quality cut of S/N > 11.2 per $60\,\mathrm{km\,s}^{-1}$ pixel to their sample, which would imply a quality cut on our sample of S/N > 4.6 per $10\,\mathrm{km\,s}^{-1}$ pixel.

While previous work noticed an increased scatter in the opacity measurements only at $z \gtrsim 5.5$, we also see evidence for increased scatter at $5.0 < z < 5.5$. We see systematic differences towards higher optical depths in our work compared to others, most strikingly in the $5.3 < z < 5.7$ bins whose excess fluctuations have been the focus of several works. However, in most redshift bins the measurements agree within 1σ-uncertainties (shown as the shaded regions in Fig. 9.4) which we determined via bootstrap resampling, the only exception being the redshift bin at $z = 5.4$ and $z = 5.6$, where our results are slightly more discrepant with previous studies.

A discrepancy between the Fan et al. (2006c) and Becker et al. (2015b) measurements in this bin was previously noted by Chardin et al. (2017), and in particular it seems that our (higher) τ_{eff} measurements are more consistent with the data from Becker et al. (2015b) than those from Fan et al. (2006c).

9.3 Simulations of the Intergalactic Medium

We would like to compare our measurements of the IGM opacity to expectations from simulations. For this purpose we use a hydrodynamical simulation which we briefly describe in section § 9.3.1. In this simulation we use a uniform UVB radiation field and thus this simulations provides a good approximation for opacity fluctuations in the IGM long after the Epoch of Reionization, when we expect to have a uniform UVB. We use two more sophisticated models to compare our measurements with more realistic conditions in the post-reionization IGM. To this end, we use two semi-numerical models with a fluctuating UVB and temperature field, which we describe in § 9.3.2. In § 9.3.3 we explain how we compute the Lyα optical depth from the skewers through the simulation box.

9.3.1 Nyx Hydrodynamical Simulation

In this work we employ a `Nyx` hydrodynamical simulation (Almgren et al., 2013) with 100 cMpc h^{-1} on a side with 4096^3 dark matter particles and gas elements on a uniform Eulerian grid, designed for precision studies of the Lyα forest (Lukić

et al., 2015). We extracted 1000 random skewers of density, temperature, and velocity along the directions of the grid axes from simulation outputs at $z =$3.0, 3.5, 4.0, 4.5, 5.0, 5.5, 6.0, and 6.5. We then computed the resulting Lyα forest spectra in 50 cMpc h^{-1} bins consistent with the scale of the IGM opacity measurements presented here. For redshift bins in between the simulation outputs, we take the closest output and re-scale the density field by $(1 + z)^3$ accordingly.

The simulation adopted the uniform UVB model of Haardt & Madau (2012), resulting in a "vanilla" IGM model which (uniformly) reionized at early times ($z_{\rm reion} >$ 10). Thus any deviations of the distribution of IGM opacity observations from that predicted in the simulation likely represent spatial inhomogeneities in the UVB (Davies & Furlanetto, 2016) or fluctuations in the thermal state of the IGM (D'Aloisio et al., 2015).

9.3.2 Semi-Numerical Models with Fluctuating UVB and Temperature Fields

We also compare our observations at $z = 5.7$ to the semi-numerical fluctuating UVB and fluctuating IGM temperature models from Davies et al. (2018a) which we describe here briefly. The Davies et al. (2018a) semi-numerical simulation consists of a cosmological volume, 780 cMpc on a side, with a 2048^3 density field computed via the Zel'dovich approximation (Zel'dovich, 1970) and dark matter halos ($M_{\rm halo} \geq 2 \times 10^9$ M$_\odot$) populated using the excursion set formalism as in Mesinger & Furlanetto (2007). Ionizing luminosities were assigned to halos by first abundance matching to the (non-ionizing) UV luminosity function (Bouwens et al., 2015) and then allowing the ratio of ionizing to non-ionizing luminosities to vary as a free parameter.

UVB fluctuations in this volume were computed on a 156^3 grid following the method of Davies & Furlanetto (2016) with a spatially-varying mean free path of ionizing photons. To construct a fluctuating IGM temperature field, the reionization redshifts of each density cell in the semi-numerical simulation were computed with 21cmFAST (Mesinger et al., 2011) and the subsequent cooling from an initial post-reionization temperature of $30,000$ K was computed via numerical integration of the IGM thermal evolution (as in Upton Sanderbeck et al. 2016). Finally, Lyα forest sightlines were then computed using the fluctuating Gunn-Peterson approximation (Weinberg et al., 1997) applied to the Zel'dovich approximation density field, with a normalization factor applied to the optical depth of each pixel to account for the approximate nature of the method.

9.3.3 Calculating the Lyα Optical Depth from Simulated Skewers

We then extract 1000 skewers through the various simulation boxes to compute the optical depths and compare the results to our measurements. Because the exact strength of the UVB radiation Γ_{UVB} is unknown, we have to re-scale the optical depth in each skewer at each pixel i, i.e. $\tau_i^{\mathrm{Ly}\alpha,\mathrm{unscaled}}$, to match the mean optical depth corresponding to the observed mean flux value $\langle F^{\mathrm{obs}} \rangle$ of our measurements, which in turn depends on the exact value of Γ_{UVB}. Hence, at each redshift we determine a scaling factor A_0 that solves the following equation:

$$\left\langle \exp\left[-\tau_i^{\mathrm{Ly}\alpha}\right] \right\rangle = \left\langle \exp\left[-A_0 \times \tau_i^{\mathrm{Ly}\alpha,\mathrm{unscaled}}\right] \right\rangle$$
$$= \left\langle F^{\mathrm{obs}} \right\rangle \tag{9.1}$$

We then average the re-scaled flux at each pixel $\left\langle \exp\left[-\tau_i^{\mathrm{Ly}\alpha}\right] \right\rangle$ within each skewer of size $50\,\mathrm{cMpc}\,h^{-1}$, and determine the 68% and 95% of the distribution of $\tau_{\mathrm{eff}}^{\mathrm{Ly}\alpha}$. This gives an estimate of the expected scatter within the Lyα optical depth measurements given a uniform UVB and IGM thermal state, that is, arising from density fluctuations alone.

9.4 Implications for the Epoch of Reionization

In order to assess the implications of our opacity measurements for the Epoch of Reionization, we compare our measurements to the opacity distributions from the Nyx hydrodynamical simulation (§ 9.3.1). The Nyx simulation was computed with a uniform UVB that we expect long after the reionization event or when assuming no signatures of an inhomogeneous reionization process. The evolution of the optical depth distributions from these simulations, with the mean fluxes matched to our measurements, are shown as the grey region in Fig. 9.2. The width of the light (68th percentile) and medium (95th percentile) grey regions corresponds to the scatter in τ_{eff} expected due to fluctuations in the underlying density field alone. The dark grey regions indicate the additional scatter expected from $\sim 20\%$ uncertainties in the quasar continuum estimation. These have been calculated by dividing the mean flux in each spectral bin by $(1 + \Delta C)$, where the continuum uncertainty ΔC was drawn randomly from a normal distribution with $\sigma = 0.2$ and $\mu = 0$, corresponding to $\sim 20\%$ uncertainties in the continuum estimate. As expected continuum uncertainties matter only very little at higher redshifts, when the mean transmitted flux is low and fluctuations between different quasar sightlines dom-
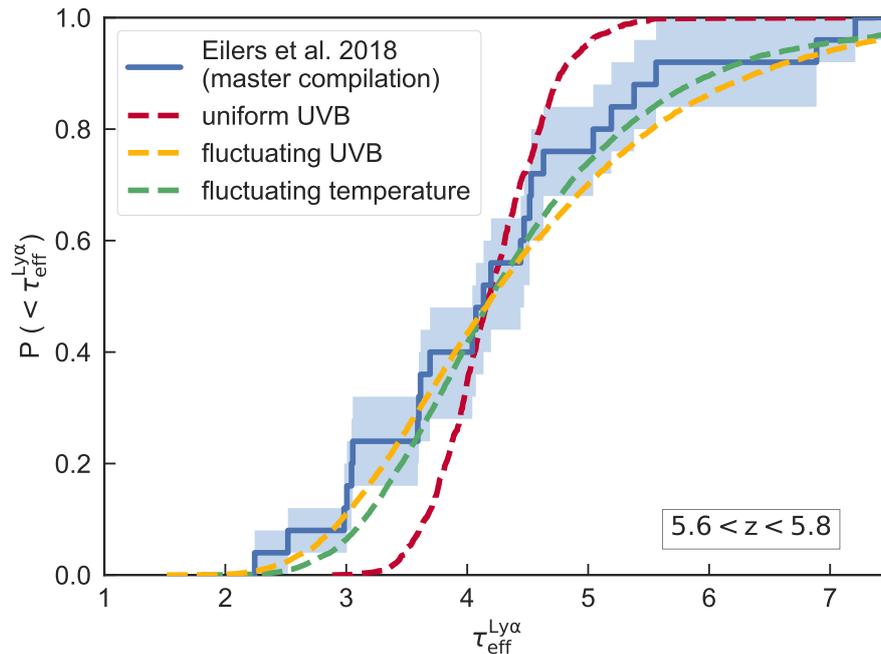
Figure 9.5 The observed CDF at $z = 5.7$ (blue curve) compared to a hydrodynam-
ical simulation with a uniform UVB model (red curve) and a semi-numerical simu-
lation with either a fluctuating UVB (yellow curve) or fluctuating post-reionization
temperature (green curve). The blue shaded regions shows the 1σ-uncertaintiy on
the CDF determined via boostrapping.

inate, and the scatter at high redshift can thus not be explained by continuum
uncertainties.

The measurements show a steep rise in τ_{eff} for $z \gtrsim 5$ and an increased scatter
in the distribution of measurements. At lower redshifts ($z \lesssim 5$) the scatter in the
observed τ_{eff} decreases rapidly and becomes consistent with the expectations from
density fluctuations alone. It is evident, however, that at high redshifts the scat-
ter in the optical depths measurements significantly exceeds the scatter expected
from density fluctuations alone, i.e. the scatter represented by our hydrodynam-
ical simulation with uniform UVB. The tiny rare flux spikes observed in the Lyα
forests of SDSS J0100 + 2802, and SDSS J1148 + 5251 at redshifts $5.8 \lesssim z \lesssim 5.9$ are
in strong contrast to the abundant transmitted flux along the sight lines towards
SDSS J1306 + 0356 or SDSS J2054 − 0005 at similar redshifts, for example. We
show the respective spectral regions exhibiting very high (upper panels), average
(middle panels) and very low (lower panels) optical depths at similar redshifts of
the aforementioned sight lines in Fig. 9.6.

The discrepancy between our measured opacity distribution and the expecta-
tion from simulations with a uniform UVB becomes even more evident in Fig. 9.4,
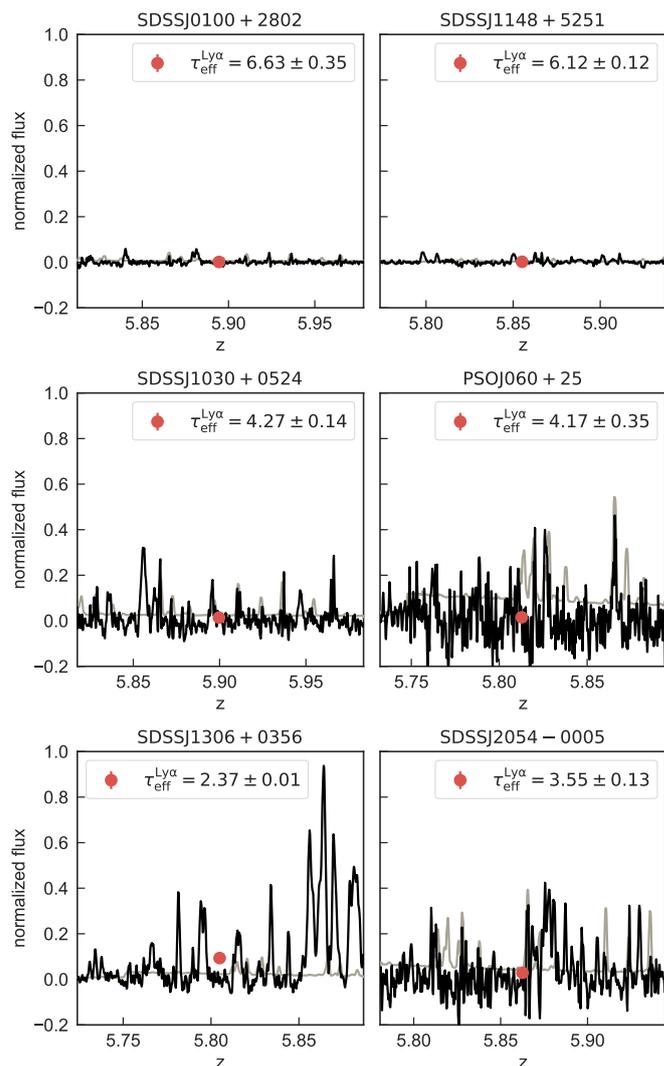
Figure 9.6 Selected spectral bins of $50\,\mathrm{cMpc}\,h^{-1}$ at a similar redshift, for which we measure $\tau_{\mathrm{eff}}^{\mathrm{Ly}\alpha}$ within the Lyα forest, along six different quasar sight lines that demonstrate the observed scatter. The sight lines in the top panels exhibit very low transmitted flux, i.e. very high optical depths, the middle panels show example spectral bins for a medium level of transmitted flux, whereas the sight lines in the bottom panels show abundant transmitted flux, i.e. very low optical depths. All spectral bins shown are at comparable redshifts. The grey curve in each panel shows the respective noise vectors and the red data points show the measurements of $\langle F^{\mathrm{obs}}\rangle$ (the errorbar is smaller than the symbol and thus not visible) with the corresponding optical depth measurements in the legend.

where we show the cumulative distributions of $\tau_{\mathrm{eff}}^{\mathrm{Ly}\alpha}$ from our master compilation in different redshift bins. The CDFs of τ_{eff} from our hydrodynamical simulation with a uniform UVB are shown as the dashed red curves, where we have rescaled the pixel optical depths (§ 9.3.3) to match the mean τ_{eff} in each redshift bin. This model with a uniform UVB is clearly not a good match to the observations. While they are more consistent with the measurements at lower redshift ($z \sim 5$), there are large discrepancies at higher redshifts ($z \gtrsim 5.6$) between the simulated and the observed CDF, a point previously noted by Fan et al. (2006c) and Becker et al. (2015b). While it may seem that the model provides a better fit in the highest redshift bin at $z \sim 6.0$, this apparent agreement is misleading, and arises due to the fact that we show limits on the optical depth in the same way as measurements, and the bin at $z \sim 6$ contains several limits.

9.4.1 Patchy Reionization Models

Multiple scenarios have been proposed to explain the increased scatter in the optical depth relative to the fluctuations one would expect from the density field of the IGM alone. One possible explanation is that the UVB is strongly fluctuating, either due to coupled variations in the mean free path of ionizing photons (Davies & Furlanetto, 2016; D'Aloisio et al., 2018) or a rare source population, such as quasars (Chardin et al., 2015, 2017). Another possibility is that the thermal state of the IGM is highly inhomogeneous (Lidz & Malloy, 2014; D'Aloisio et al., 2015). Such fluctuations can arise as a result of an extended and patchy reionization process, where different regions in the universe were reionized (and simultaneously photoheated) at different redshifts z_{reion}. The regions that reionized earlier would have had time to cool down while regions that were reionized at a later time would still be relatively hot.

In Fig. 9.5 we compare our measurements at $5.6 < z < 5.8$ to the semi-numerical models with a fluctuating UVB and fluctuating temperature field (see § 9.3.2), where we have re-scaled the opacities in the Lyα forest skewers from that work to match the mean τ_{eff} we have measured in this bin. Note that while previously the CDF of optical depths at this redshift bin containing the strong outliers in the opacity measurements in the GP trough along the sightline of ULASJ0148 + 0600 (Becker et al., 2015b), was very challenging to reproduce in simulations (D'Aloisio et al., 2015; Davies & Furlanetto, 2016; Davies et al., 2018a; Chardin et al., 2017; Keating et al., 2018), these outliers are now easier to explain because the mean τ_{eff} in our measurements is substantially higher than the Fan et al. (2006c)+Becker et al. (2015b) compilation, and hence these data points do not represent such strong deviations from the mean of the distribution anymore. This first comparison of our measurements to the two semi-numerical models with a fluctuating UVB and a fluctuating thermal state of the IGM shows that both models can reproduce the observations. A more detailed comparison to these models will be part of future work.

9.5 Summary

For a sample of 23 quasar spectra, that do not show broad absorption line features and have good quality data (i.e. S/N > 7), we measure the IGM opacity by means of the effective optical depth of the Lyα forest in bins of 50 cMpc h^{-1} covering a redshift range of $4.0 \lesssim z \lesssim 6.5$. Our results are in qualitative agreement with previous studies (Fan et al., 2006c; Becker et al., 2015b; Bosman et al., 2018), showing a

steep rise in opacity and increased scatter within the measurements at high redshift. However, while previous work noticed an increased scatter at $z \gtrsim 5.5$, we also see evidence for increased scatter at $5.0 < z < 5.5$. A detailed comparison in the optical depth in several redshift bins, shows systematic differences towards higher optical depths in our work compared to others, most strikingly at $5.3 < z < 5.7$. The discrepancies, however, between our measurements and previous work are mostly within the $\sim 1\sigma$ uncertainties, which we determined via bootstrap resampling.

Our work improves upon previous studies in several aspects. We considered possible contamination due to intervening low-ion metal absorption systems such as DLAs that have previously been ignored and carefully masked all regions in the Lyα forest that are affected by these high H I column density absorption systems. We also corrected for small offsets in the zero-level of the quasar spectra, introduced presumably by improper sky subtraction of a few individual exposures. Finally and most importantly, we considered a very homogeneously reduced data sample which minimizes systematic effects due to the use of different telescopes and detectors, or data reduction pipelines. We present a master compilation sample including mainly our newly analyzed sample with the exception of the sightline of ULAS J0148 + 0600 taken from Becker et al. (2015b), who has a larger sensitivity in the prominent GP trough due to the higher signal-to-noise ratio spectrum observed with X-Shooter on the Very Large Telescope (VLT).

We compare our measurements to a large-volume hydrodynamical simulation with a uniform UVB. As noted previously by Fan et al. (2006c) and Becker et al. (2015b), we find that the spread in observed $\tau_{\rm eff}$ cannot be explained by fluctuations of the underlying density field alone, and thus our results support an inhomogeneous reionization scenario. Whether temperature fluctuations in the IGM, a fluctuating UVB or a combination of both can best explain this increased scatter in opacity, remains an open question. A preliminary comparison of our measurements to semi-numerical simulations of UVB and IGM temperature fluctuations shows good agreement for both scenarios.

This work presents a crucial ingredient in constraining the end-stages of the Epoch of Reionization at $5.0 \lesssim z \lesssim 6.0$, when the physical conditions of the post-reionization IGM can be directly measured via absorption spectroscopy of high redshift quasars. The past several years have seen an impressive fivefold increase in the number of $z > 6$ quasars from deep wide-field optical and infrared surveys, which are enabling precise measurements of the Lyα forest absorption at $5 < z < 6.5$ (Becker et al., 2015b; Gnedin et al., 2017; Davies et al., 2018a). The requirement that reionization models reproduce these high-precision measurements provides

an important low redshift anchor point which all models must reproduce, and can dramatically narrow the family of viable reionization models. Statistical anaylses of the Lyα forest, such as measurements of the power spectrum (Oñorbe et al., 2017b; D'Aloisio et al., 2018) or the probability distribution function (PDF) of the IGM opacity (Davies et al., 2018b) set further constraints on the reionization process, allowing us to develop accurate models about the early evolutionary stages of our universe.

Possible Anomaly in the Lyβ Forest Opacity of the IGM

In this chapter we explore a different tracer of the intergalactic gas, which is co-spatial with the Lyα forest, namely the Lyβ forest. Due to the approximately 5 times lower oscillator strength of the Lyβ transition, the Lyβ forest is more sensitive to gas with a higher neutral fraction and thus provides more stringent constraints on the IGM ionization state (Lidz et al., 2002; Songaila, 2004; Fan et al., 2006c). Furthermore, the ratio of the Lyα to Lyβ forest opacity is sensitive to the thermal state of the IGM (Oh & Furlanetto, 2005), since the neutral fraction of a gas in photoionization equilibrium is proportional to the gas density (see Eqn. (6.2)), and the temperature of the intergalactic gas also correlates with its density (see Eqn. (6.1)).

While several authors have measured the evolution of Lyα and Lyβ opacity and implications for the neutral fraction of the IGM (Lidz et al., 2002; Songaila, 2004; Fan et al., 2006c), the correspondence between Lyα and Lyβ opacities has not been studied in detail. Here, we address this connection and measure the IGM opacity in both the Lyα and the Lyβ forest towards the end of the reionization epoch between $5.5 \leq z \leq 6.1$ along 19 quasar sightlines in our data set which have S/N $\gtrsim 10$ per pixel and do not show any BAL signatures. We present the measurements for both the Lyα and Lyβ forest in § 10.1. In § 10.2 we compare these measurements to different models of physical conditions in the post-reionization IGM based on hydrodynamical simulations. We discuss the implications of our results on the Epoch of Reionization in § 10.3.

10.1 The Effective Optical Depth in the Lyα and Lyβ Forest

As in the previous chapter we estimate the IGM opacity by means of the effective optical depth, which is shown in Eqn. (6.6). In this study, we choose the size of the spectral bins over which we average the normalized flux to be 40 cMpc, instead of the previously applied bin size of 50 cMpc h^{-1} (Becker et al., 2015b; Eilers et al., 2018a; Bosman et al., 2018), such that the Lyβ forest region (i.e. between the wavelengths of Lyβ and Lyγ) is better sampled. Note that whenever the average flux $\langle F \rangle$ is negative or detected with less than 2σ significance, we adopt a lower limit on the optical depth at the 2σ level.

For a quasar at $z_{\rm em} \sim 6$ the Lyα and Lyβ forests lie between the observed wavelengths $7180 - 8510$ Å and $6808 - 7180$ Å, respectively. However, we do not conduct the IGM opacity measurements within the whole wavelength region, but exclude the proximity zones around each quasar, and choose the rest-frame wavelength 1030 Å and 975 Å as the minimum wavelengths for our measurements in the Lyα and Lyβ forest, respectively. Additionally, as described in the previous chapter, we mask the spectral regions around DLAs, in order to avoid any biases in the estimation of the IGM opacity.

Within the observed wavelength range of the Lyβ forest we also observe Lyα forest absorption from the foreground IGM at lower redshift, i.e.

$$z_{\rm fg} = \left((1+z) \frac{\lambda_{\rm Ly\alpha}}{\lambda_{\rm Ly\beta}} \right) - 1. \tag{10.1}$$

The redshift range of the foreground Lyα forest absorption within the Lyβ forest is $z_{\rm fg} \approx 4.6 - 4.9$ for a quasar at $z_{\rm em} \sim 6$. For this study we calculate the effective optical depth in the Lyα forest $\tau_{\rm eff}^{\rm Ly\alpha}$ as well as the *observed* effective optical depth in the Lyβ forest $\tau_{\rm eff}^{\rm Ly\beta,obs}$, which includes the absorption from the foreground Lyα forest. Hence, we do not attempt to "correct" the Lyβ opacity measurements by subtracting the foreground Lyα absorption, but rather analyze the observed Lyβ opacity, which is the sum of the pure Lyβ optical depth and the Lyα forest opacity in the foreground IGM. We measure both effective optical depths within the Lyα and the Lyβ forest in discrete spectral bins around the same central redshift $z_{\rm abs}$. Our chosen bin size of 40 cMpc allows us to obtain two estimates along each quasar sight line, in the absence of any masked spectral bins due to intervening DLAs[1].

---

[1]Note that the masking of bins containing a DLA removes all bins within the Lyβ forest of two quasar sightlines, i.e. SDSS J0100 + 2802 and SDSS J1148 + 5251
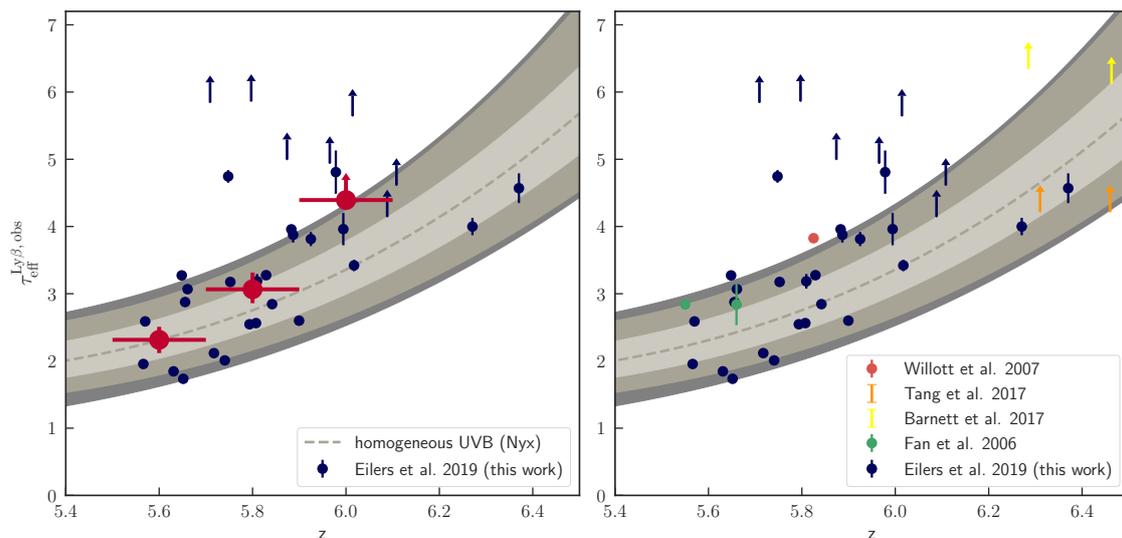
Figure 10.1 Evolution of the observed Lyβ optical depth with redshift. The dark blue data points in both panels show our measurements of $\tau_{\mathrm{eff}}^{\mathrm{Ly}\beta,\mathrm{obs}}$. *Left panel:* The large red data points show the opacity evolution evolution averaged over bins of $\Delta z = 0.2$, with uncertainties determined via bootstrapping. The grey underlying region shows the predicted redshift evolution from hydrodynamical simulations assuming a uniform UVB model. We have simulation outputs in steps of $\Delta z = 0.5$ and use a cubic spline function to interpolate the shaded regions between the redshift outputs. The light and medium grey shaded regions indicate the 68th and 95th percentile of the scatter expected from density fluctuations in the simulations, whereas the dark grey regions show any additional scatter due to ~ 20% continuum uncertainties. *Right panel:* Compilation of all Lyβ opacity measurements found in the literature along quasar sightlines that are not in our data sample. Note, however, that the chosen spectral bin sizes here vary between different analyses.

Our resulting measurements of the effective optical depth in both the Lyα and the Lyβ forests along all quasar sight lines in our data sample are listed in Table 10.1. The redshift evolution of the observed effective optical depth in the Lyβ forest is shown in Fig. 10.1. The left panel shows also the mean $\tau_{\mathrm{eff}}^{\mathrm{Ly}\beta,\mathrm{obs}}$ in bins of $\Delta z = 0.2$, whereas the right panel shows a compilation of all Lyβ measurements from the literature (Fan et al., 2006c; Willott et al., 2007; Tang et al., 2017; Barnett et al., 2017). We find a sharp increase in the observed Lyβ optical depth between $z \sim 5.5$ and $z \sim 6$, which is in agreement with previous results (Lidz et al., 2002; Songaila, 2004; Fan et al., 2006c). We show the transmitted flux in the Lyα forest and the corresponding region at the same absorption redshift in the Lyβ forest within several spectral bins along four different quasar sightlines in Fig. 10.2.

Table 10.1 Mean flux measurements in the Lyα and Lyβ forest.

object	z_{em}	z_{start}	z_{abs}	z_{end}	$\langle F^{Ly\alpha}\rangle$	$\langle F^{Ly\beta}\rangle$
SDSSJ0002+2550	5.82	5.699	5.656	5.613	0.0728 ± 0.0005	0.0563 ± 0.0005
		5.613	5.570	5.528	0.0245 ± 0.0006	0.0752 ± 0.0005
SDSSJ0005-0006	5.844	5.761	5.717	5.674	0.0582 ± 0.0022	0.1205 ± 0.0066
		5.674	5.631	5.588	0.0484 ± 0.0027	0.1580 ± 0.0067
CFHQSJ0050+3445	6.253	6.136	6.088	6.041	0.0027 ± 0.0025	-0.0018 ± 0.0079
		6.041	5.994	5.948	0.0101 ± 0.0031	0.0190 ± 0.0046
ULASJ0148+0600	5.98	5.842	5.797	5.753	-0.0022 ± 0.0032	-0.0032 ± 0.0014
		5.753	5.709	5.666	0.0028 ± 0.0021	-0.0059 ± 0.0014
PSOJ036+03	6.5412	6.421	6.370	6.320	0.0010 ± 0.0021	0.0104 ± 0.0023
		6.320	6.271	6.222	-0.0034 ± 0.0021	0.0183 ± 0.0024
PSOJ060+25	6.18	6.064	6.017	5.971	0.0623 ± 0.0039	0.0327 ± 0.0028
		5.971	5.925	5.879	0.0163 ± 0.0081	0.0221 ± 0.0023
SDSSJ0836+0054	5.81	5.695	5.652	5.609	0.1323 ± 0.0003	0.1765 ± 0.0006
		5.609	5.566	5.524	0.0887 ± 0.0005	0.1416 ± 0.0005
SDSSJ0840+5624	5.8441	5.692	5.648	5.606	0.0124 ± 0.0013	0.0379 ± 0.0014
SDSSJ1030+0524	6.309	6.156	6.108	6.061	0.0097 ± 0.0020	-0.0136 ± 0.0050
		6.061	6.014	5.968	0.0018 ± 0.0022	-0.0156 ± 0.0018
SDSSJ1137+3549	6.03	5.874	5.829	5.785	0.0126 ± 0.0040	0.0378 ± 0.0021
		5.785	5.741	5.697	0.0068 ± 0.0031	0.1342 ± 0.0022
SDSSJ1306+0356	6.016	5.887	5.842	5.797	0.1467 ± 0.0022	0.0581 ± 0.0010
		5.797	5.752	5.709	0.0486 ± 0.0015	0.0418 ± 0.0010
ULASJ1319+0950	6.133	6.025	5.978	5.932	-0.0021 ± 0.0064	0.0082 ± 0.0026
		5.932	5.887	5.841	0.0265 ± 0.0089	0.0207 ± 0.0024
SDSSJ1411+1217	5.904	5.792	5.748	5.704	0.0131 ± 0.0012	0.0087 ± 0.0008
		5.704	5.661	5.618	0.0362 ± 0.0011	0.0465 ± 0.0008
SDSSJ1602+4228	6.09	5.929	5.883	5.838	0.0136 ± 0.0040	0.0191 ± 0.0013
		5.838	5.794	5.749	0.0604 ± 0.0043	0.0784 ± 0.0014
SDSSJ1630+4012	6.065	5.852	5.807	5.763	0.0366 ± 0.0071	0.0771 ± 0.0028
SDSSJ2054-0005	6.0391	5.946	5.900	5.855	0.0519 ± 0.0055	0.0743 ± 0.0047
		5.855	5.810	5.765	0.0085 ± 0.0059	0.0414 ± 0.0045
SDSSJ2315-0023	6.117	6.012	5.965	5.919	-0.0073 ± 0.0040	-0.0068 ± 0.0036
		5.919	5.874	5.829	-0.0135 ± 0.0073	-0.0038 ± 0.0034

Notes. The different columns show the name of the object and its emission redshift z_{em}, the start of the redshift bin z_{start}, the mean redshift of each bin z_{abs}, and the end of each bin z_{end}, as well as the measured mean flux of the continuum normalized spectra in the Lyα and Lyβ forest.

10.2 Models of the Post-Reionization IGM

As in the previous chapters, we would like to compare our measurements of the IGM opacity to models of the post-reionization IGM. Hence we run several hydrodynamical simulations: first, we assume a uniform UVB and several variations of the thermal state of the IGM (§ 10.2.1). Second, we model a spatially fluctuating reionization process with an inhomogeneous UVB, or variations in the temperature field of the IGM (§ 10.2.2). We will then calculate the optical depths along skewers through the simulation boxes (§ 10.2.3) and forward-model the spectral noise of the data (§ 10.2.4).

10.2.1 Modeling Variations in the Thermal State of the IGM

We introduced the cosmological hydrodynamical simulation Nyx (Almgren et al., 2013) already in the previous chapter in § 9.3.1. Additionally, for the study of the Lyβ forest optical depth we use simulation outputs at lower redshifts to model the contamination of the Lyβ forest opacity by foreground Lyα absorption (see § 10.2.3). For redshifts in between the simulation outputs, we take the closest output and rescale the density field by $(1+z)^3$ accordingly.

In accordance with the standard paradigm for the thermal state of the photoionized IGM after reionization events, we expect that the majority of the optically thin gas responsible for the absorption in the Lyα and Lyβ forests follows a tight temperature-density relation (Hui & Gnedin, 1997), which we introduced in § 6.2 in Eqn. (6.1). In our simulations we vary the slope of the temperature-density relation in post-processing, imposing values of $\gamma = 0.0$, which indicates an inversion of the thermal state of the IGM, $\gamma = 1.0$, representing an isothermal state of the IGM, and $\gamma = 1.5$, which is the fiducial value for the IGM long after any reionization events (Hui & Gnedin, 1997; McQuinn & Upton Sanderbeck, 2016).

The grey bands underlying the measurements in Fig. 10.1 show the expected redshift evolution of the Lyβ opacity from this hydrodynamical simulation with a uniform UVB and $\gamma = 1.5$. The light and medium grey bands show the expected 1σ and 2σ scatter, where as the thin dark grey band presents any additional scatter expected from $\sim 20\%$ uncertainties in the quasar continuum estimate, which only have a small effect (Becker et al., 2015b; Eilers et al., 2017b). As noticed in multiple previous studies the observed scatter is larger than expected from a uniform UVB model, which gave rise to models with spatial fluctuations in the UVB or the underlying temperature field (Fan et al., 2006c; Becker et al., 2015b; Eilers et al., 2018a; Bosman et al., 2018).

79
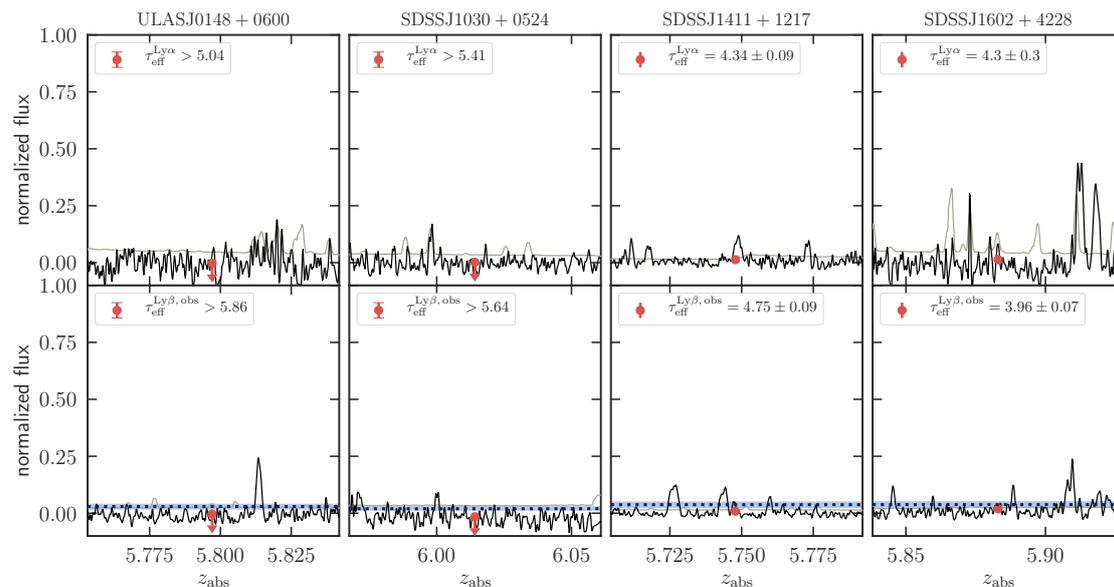
Figure 10.2 Selected spectral bins along four different quasar sightlines showing the transmitted flux in the Lyα forest (top panels) and the Lyβ forest (lower panels) at the same absorption redshift z_{abs}. The red data points indicate the measured mean fluxes $\langle F \rangle$ in the respective bins and the corresponding optical depths are indicated in the legend. Note that the errorbars on the mean fluxes are smaller than the symbols. The blue bands in the lower panels indicate the expected mean flux in the Lyβ forest from hydrodynamical simulations with a fluctuating UVB (see § 10.2.2).

10.2.2 Fluctuating Reionization Models

Thus we also compare our measurements to predictions of the Lyα and Lyβ forest opacity from published models with a spatially inhomogeneous UVB by Davies et al. (2018b). The model consists of UVB fluctuations from a $(400 \text{ cMpc})^3$ semi-numerical model with an average mean free path of ionizing photons of 25 cMpc, computed with a spatially-varying mean free path following Davies & Furlanetto (2016), which we apply to skewers from our Nyx simulations. This approach modestly overestimates the effect of UVB fluctuations on the IGM opacity because the UVB-density anti-correlation (Davies & Furlanetto, 2016; Davies et al., 2018a) is lost. While small-scale correlations between the UVB and the density field may alter the relationship between Lyα and Lyβ forest opacity (Oh & Furlanetto, 2005), the UVB fluctuations in the Davies et al. (2018a) model manifest on scales comparable to the size of our 40 cMpc bins, so we do not expect a substantial effect.

For modeling a fluctuating temperature field of the IGM we employ the IR-A model of Oñorbe et al. (2018). An inhomogeneous reionization history was computed for a 40 Mpc/h volume using the semi-numerical reionization code 21cmFAST

(Mesinger et al., 2011), which was then used to apply time-dependent reionization heating ($\Delta T = 25,000$ K) to a 2048^3 `Nyx` hydrodynamical simulation of the same cosmological volume.

10.2.3 Calculating the Lyβ Optical Depth from Simulated Skewers

We extract 1000 skewers from each model and compute the mean fluxes and optical depths along discrete spectral bins of the same size as our measurements, i.e. 40 cMpc. Since the observed flux depends on the uncertain mean photoionization rate from the UVB Γ_{UVB}, we re-scale the optical depth at each pixel to match the mean optical depth corresponding to the observed mean flux value $\langle F_{\mathrm{Ly}\alpha}^{\mathrm{obs}} \rangle$ of our measurements according to Eqn. (9.1).

We obtain the *pure* Lyβ optical depth $\tau_i^{\mathrm{Ly}\beta}$ at each pixel from $\tau_i^{\mathrm{Ly}\alpha}$ by scaling according to the oscillator strengths, $f_{\mathrm{Ly}\alpha}$ and $f_{\mathrm{Ly}\beta}$, of their transitions and the respective wavelengths, i.e.

$$\tau_i^{\mathrm{Ly}\beta} = \frac{f_{\mathrm{Ly}\beta}}{f_{\mathrm{Ly}\alpha}} \frac{\lambda_{\mathrm{Ly}\beta}}{\lambda_{\mathrm{Ly}\alpha}} \tau_i^{\mathrm{Ly}\alpha} \tag{10.2}$$

with $f_{\mathrm{Ly}\alpha} = 0.41641$ and $f_{\mathrm{Ly}\beta} = 0.079142$ (Table 4 in Wiese & Fuhr, 2009).

However, the *observed* Lyβ effective optical depth $\tau_{\mathrm{eff}}^{\mathrm{Ly}\beta,\mathrm{obs}}$ will be higher than the pure Lyβ optical depth due the additional absorption from the foreground Lyα forest. Thus we create additional skewers along *different* sight lines with Lyα forest absorption at the foreground redshift $\tau_i^{\mathrm{Ly}\alpha,\mathrm{fg}}$, using lower redshift outputs from the `Nyx` simulation. These lower redshift skewers are also re-scaled to match the observed mean flux at z_{fg}, for which we use the formula by Oñorbe et al. (2017a) to fit the literature measurements of the optical depth by Faucher-Giguère et al. (2008).

At each pixel we now sum the pure Lyβ forest optical depth with the foreground Lyα optical depth to obtain the flux one would observe in the Lyβ forest, i.e.

$$F_i^{\mathrm{Ly}\beta,\mathrm{obs}} = \exp\left[-(\tau_i^{\mathrm{Ly}\beta} + \tau_i^{\mathrm{Ly}\alpha,\mathrm{fg}}) \right], \tag{10.3}$$

and then calculate the observed effective Lyβ optical depth in each spectral bin by averaging the flux of all pixels within the bin, $\tau_{\mathrm{eff}}^{\mathrm{Ly}\beta,\mathrm{obs}} = -\ln \langle F_i^{\mathrm{Ly}\beta,\mathrm{obs}} \rangle$.

10.2.4 Forward-Modeling of Spectral Noise

The spectral noise in the data influences the opacity measurements, such that whenever the averaged flux is detected with less than 2σ significance, we adopt

a lower limit on the optical depth measurement (see § 10.1). To mimic the effects of the noisy data, we draw randomly for each mean flux value $\langle F_{\rm sim} \rangle$ along a simulated skewer an uncertainty $\sigma_{\langle F_{\rm data} \rangle}$ from the measurements. We then apply the same criterion as for the real data and adopt a measurement only if $\langle F_{\rm sim} \rangle \geq 2\sigma_{\langle F_{\rm data} \rangle}$, or a lower limit at $2\sigma_{\langle F_{\rm data} \rangle}$ otherwise. This reduces the sensitivity to very high opacities in the simulated skewers, which we do not have in noisy data.

10.3 Constraints on the Thermal and Ionization State of the IGM

We compare our measurements of $\tau_{\rm eff}^{\rm Ly\alpha}$ and $\tau_{\rm eff}^{\rm Ly\beta,obs}$ to predictions from the various models of the post-reionization IGM in three different redshift bins in Fig. 10.3. The top panels show models with a uniform UVB, but varying slopes of the temperature-density relation of the IGM, whereas the middle and bottom panels show simulations with spatial fluctuations in the UVB and the temperature field, respectively. The dotted contours in the middle and lower panels show the additional scatter when accounting for uncertainties in the continuum estimate of 20%, which we did not show in the top panels for better readability. Note that we did not yet forward-model the spectral noise in the data to the simulations for this qualitative comparison.

At lower redshift, i.e. $5.5. < z < 5.7$, all models seem to reproduce the observations very well. However, while the observations in the Lyα forest opacity at higher redhifts, which have been subject of considerable modeling efforts in recent years (Becker et al., 2015b; D'Aloisio et al., 2015; Chardin et al., 2015; Davies & Furlanetto, 2016), are well captured by the simulations, the $\tau_{\rm eff}^{\rm Ly\beta,obs}$ measurements show a significantly larger scatter than the models predict, and also their mean optical depth is higher than predicted in the highest redshift bin at $z \sim 6$. Although it is still not fully capturing the measurements, the model with a uniform UVB but an inverted temperature-density relation, i.e. $\gamma = 0.0$ matches the observations best, because it predicts a higher mean Lyβ optical depth than all other models.

In Fig. 10.4 we show the CDFs of the opacity measurements in the same redshift bins, as well as the predicted CDFs from the post-reionization model including a fluctuating UVB. Note that we show *all* $\tau_{\rm eff}^{\rm Ly\alpha}$ measurements along the complete Lyα forest rather than only the measurements which have a corresponding measurement of $\tau_{\rm eff}^{\rm Ly\beta,obs}$ at the same absorption redshift, which are reported in Table 10.1.
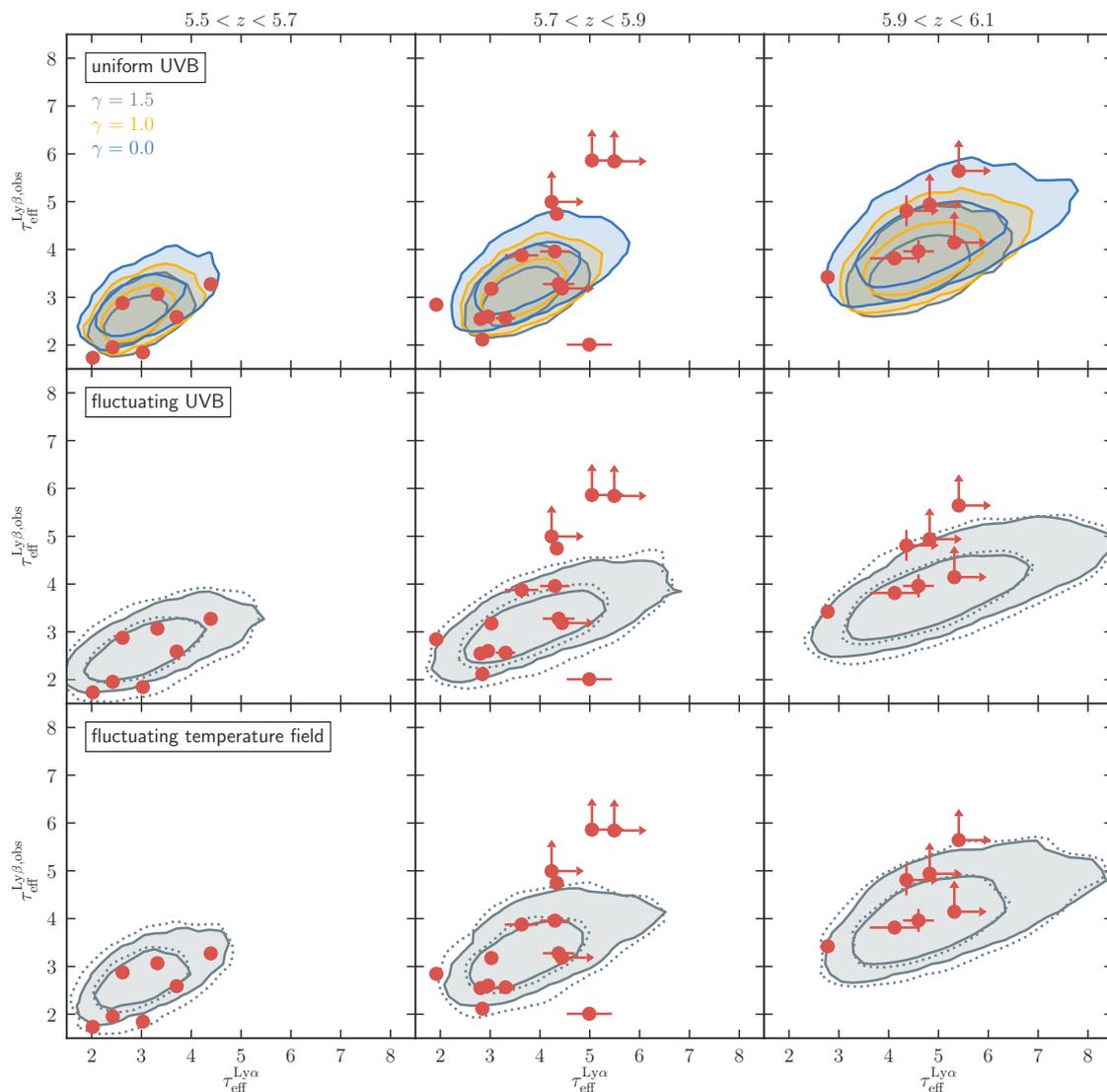
Figure 10.3 Comparison of our Lyα and Lyβ opacity measurements shown as the red data points in different redshift bins, i.e. $5.5 < z < 5.7$ (left), $5.7 < z < 5.9$ (middle), and $5.9 < z < 6.1$ (right), to predictions from several hydrodynamical simulations of the post-reionization IGM. The contours in the top panels show the prediction from simulations with uniform UVB and different slopes of the temperature-density relation of the IGM, whereas the middle and bottom panels show predictions from models with a fluctuating UVB or a fluctuating temperature field, respectively. Inner and outer contours show the 68th and 95th percentile of the distribution. The dotted contours show the respective distributions including $\sim 20\%$ continuum uncertainties.

The dotted black curves illustrate the predicted CDFs from the post-reionization model with spatial UVB fluctuations, which is believed to best reproduce the opacity measurements in the Lyα forest (Becker et al., 2015b; Davies et al., 2018a). However, in these models an infinite S/N ratio of the data is assumed, and, since
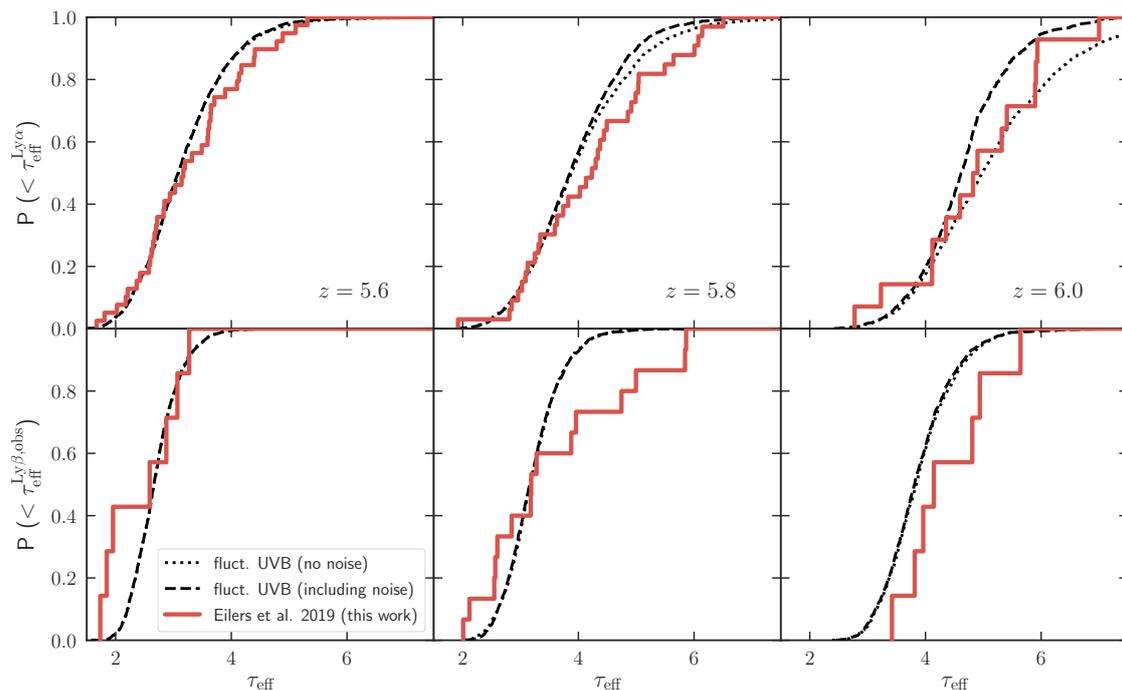
Figure 10.4 CDFs of $\tau_{\text{eff}}^{\text{Ly}\alpha}$ (top panels) and $\tau_{\text{eff}}^{\text{Ly}\beta,\text{obs}}$ (lower panels) in the same redshift bins as in Fig. 10.3. Our measurements are shown in red, the noise-free predictions from the a reionization model with a fluctuating UVB is shown as the black dotted line, whereas the black dashed curves show the same model, now including forward-modeled spectral noise. The models are scaled to match the 25th percentile of the observations. Note that the top panels do not only contain the measurements of $\tau_{\text{eff}}^{\text{Ly}\alpha}$ that have a corresponding $\tau_{\text{eff}}^{\text{Ly}\beta,\text{obs}}$ measurement at the same redshift, but rather all measurements within the Lyα forest along all 19 quasar sight lines.

our data is noisy, it is not sensitive to the very high opacity regions. Thus we forward-model the spectral noise to the simulated skewers according to the procedure described in § 10.2.4, which limits the sensitivity to regions with high optical depths and effectively reduces the amount of scatter in the models. As a result the post-reionization model does not predict enough spatial fluctuations once we account for the spectral noise in the data, which is shown as the black dashed curves in Fig. 10.4. Hence, we conclude that the forward-modeling of the spectral noise properties to the models is essential to conduct a fair comparison, and our results suggest that stronger fluctuations, i.e. a shorter mean free path of the photons, than previously assumed are needed to reproduce the measurements correctly. At lower opacities and higher mean fluxes the effect of the forward-modeling of the noise is less influential, since most measurements are a $> 2\sigma$-detection rather than a lower limit.

Additionally, we find the same offset in the mean flux within the Lyβ forest between the observations and the models that we already observed in Fig. 10.3. After fine-tuning the models to reproduce the observed mean flux in the Lyα forest by means of Eqn. (9.1), the mean flux in the Lyβ forest is clearly overestimated, and hence the models predict a much lower opacity than we observe. This difference does not exist at lower redshift, i.e. $5.5 < z < 5.7$, but becomes more pronounced in the two higher redshift bins, when approaching the Epoch of Reionization.

It is worth pointing out that in the two highest redshift bins, there is a difference of $\Delta\tau_{\text{eff}}^{\text{Ly}\beta} \approx 1$ at the high opacity end, which corresponds to a factor of ~ 2.5 between the observed mean Lyβ forest optical depth and the hydrodynamical simulations. Hence systematic uncertainties in the data that could be introduced due to the quasar continuum estimation or data reduction issues, will have to change the observed flux by $\gtrsim 250\%$ to mitigate this offset, which is highly unlikely.

In this chapter we presented new measurements of the IGM opacity in the Lyα as well as the Lyβ forest along 19 quasar sightlines. The Lyβ forest is more sensitive to higher density regions in the IGM due to the lower oscillator strength of its transition, and can thus provide more stringent constraints on post-reionization models of the IGM. A comparison to several hydrodynamical simulations revealed two main results: First, it is essential to account for the spectral noise in the data to model the observations correctly. When forward-modeling the spectral noise we find evidence for stronger spatial fluctuations in the post-reionization IGM than previously assumed. Second, there is a significant offset in the observed mean flux in the Lyβ forest compared to predictions from several models of the post-reionization IGM, which overestimate the observed flux by a factor of ~ 2.5 at high redshifts. Simulations with a uniform UVB and a highly inverted temperature-density relation of the IGM, i.e. $\gamma = 0.0$, seem to reproduce the observed ratio of Lyα and Lyβ opacities best, but are still not fully capturing the measurements. Further modeling will be required to explain the observed opacity both in the Lyα and the Lyβ forest, which will hopefully lead to new insights regarding the morphology, timing, and evolution of the reionization process.

Quasar Proximity Zones

In this chapter we measure the sizes of proximity zones for our data set of high redshift quasars that we presented in Chapter 8. Proximity zones are the regions of enhanced transmitted flux in the vicinity of the quasars, ionized by the quasars' own radiation (see § 6.5). It has been argued that the redshift evolution of the proximity zone sizes reflects the evolution of the neutral gas fraction of the IGM during the Epoch of Reionization (Fan et al., 2006c; Carilli et al., 2010). In this chapter we will show that the sizes of proximity zones are independent on the ionization state of the IGM but rather provide the unique opportunity to estimates the age of individual quasars.

The dozens of $z \sim 6$ quasars that have been uncovered over the past decade (see Bañados et al., 2016, for a recent compilation) from wide field surveys results in many new quasars for proximity zone measurements. The statistical power of these data alone motivates revisiting this type of analysis to further understand the resulting constraints on reionization and the quasar emission timescales. In this chapter we re-investigate the evolution of the quasar proximity zone sizes at $z \geq 5.77$ with a homogeneous analysis of a significantly enlarged sample of 31 quasar spectra (we exclude the three BAL quasars in our data set) with higher quality data, including updated and more precise redshift measurements from CO and [C II] line observations, and consistently measured absolute magnitudes. We will then compare our analysis with state-of-the-art radiative transfer simulations in order to better understand and interpret our measurements.

11.1 Measuring the Sizes of Proximity Zones

In order to estimate the proximity zone sizes for each quasar we adopt the standard definition used in the literature (e.g. Fan et al., 2006c; Willott et al., 2007, 2010; Carilli et al., 2010). Namely, we take the continuum normalized quasar spectra and smooth them with a 20 Å-wide (observed frame) boxcar function. This smoothing scale corresponds to ≈ 1.0 proper Mpc (pMpc) or a ≈ 705 km/s window at $z = 6$. We define the proximity zone size as the distance to the first of three consecutive pixels[1] on the blue side of the Lyα emission line that show a drop of the smoothed flux below the 10% level. This is demonstrated in Fig. 11.1 and Fig. 11.2 for subsets of bright ($-27.5 \leq M_{1450} \leq -26.5$) and faint ($-25.8 \leq M_{1450} \leq -24.5$) quasars from our data sample, respectively. The depicted quasars all cover a similar range in luminosity, but nevertheless show a wide range of proximity zone sizes, between 0.8 pMpc $\lesssim R_p \lesssim 7.1$ pMpc for bright quasars and 0.5 pMpc $\lesssim R_p \lesssim 4.0$ pMpc for fainter quasars. All measurements of the proximity zone sizes R_p for our data set are listed in Table 11.1.

11.2 Radiative Transfer Simulations

To interpret our measurements of R_p, we run a series of radiative transfer simulations of the effect of quasar ionizing photons on the IGM along the line of sight similar to those performed by Bolton & Haehnelt (2007b). We apply the one-dimensional ionizing radiative transfer code from Davies et al. (2016) to skewers from a 100 Mpc h^{-1} Eulerian hydrodynamical simulation run with the Nyx code (Almgren et al., 2013; Lukić et al., 2015) at $z = 6$ with 4096^3 baryonic (Eulerian) grid elements and dark matter particles. The radiative transfer computes the time-dependent ionization and recombination of six species (e^-, H I, H II, He I, He II, He III) as well as the associated photoionization heating and cooling by various processes including adiabatic cooling due to the expansion of the universe and inverse Compton cooling off CMB photons (see Davies et al., 2016, for details). We use 900 skewers of density, temperature, and peculiar velocity drawn along the x, y, and z grid axes from the centers of the 150 most massive dark matter halos in the Nyx simulation, corresponding to halo masses $M_h \gtrsim 4 \times 10^{11}$ M$_\odot$. In detail we find that starting the skewers from more or less massive halos has a negligible impact on the resulting proximity zone sizes, in agreement with Bolton & Haehnelt (2007b) and Keating et al. (2015).

---

[1]Note that we checked whether it changes the measurements of the proximity zone sizes when requiring up to ten pixels to be below the 10% level, but we do not observe a significant difference.

Given the relatively coarse resolution of the Nyx simulation ($25 \, \text{kpc} \, h^{-1}$) and the lack of metal-line cooling, star formation, and feedback processes, we do not expect to resolve or accurately model dense gas inside of galaxies. For this reason we have removed one object in our sample (SDSS J0840 + 5624; see Appendix D) which exhibits strong associated metal line absorption at the quasar redshift that likely arises from dense gas in the vicinity of the quasar or from its host halo.

Quasar absolute magnitudes were converted to specific luminosity at the hydrogen ionizing edge ($\nu = \nu_{\text{HI}}$) using the Lusso et al. (2015) spectral template, and for consistency with previous studies we assume that the spectrum at $\nu > \nu_{\text{HI}}$ behaves as a power law[2] with $L_\nu \propto \nu^{-1.7}$. We assume that the quasar turns on abruptly and emits at constant luminosity for its entire age t_{Q}, i.e. a so-called "light bulb" model. For the quasar age we assume a fiducial value of $t_{\text{Q}} = 10^{7.5}$ yr, but we later investigate the dependence of R_p on t_{Q} in § 11.4.3.

We consider two initial conditions for the ionization state of gas in the simulation: either the gas is initially highly ionized by a uniform ionizing background or the gas is initially completely neutral. In the highly ionized case, we add ionizing radiation due to the UVB to each cell leading to an ionization rate $\Gamma_{\text{UVB}} = 2 \times 10^{-13}$ s^{-1}, consistent with observations of Lyα forest opacity at $z \sim 6$ (e.g. Wyithe & Bolton, 2011) and resulting in a neutral fraction of $x_{\text{HI}} = 1.5 \times 10^{-4}$, with a spectrum characteristic of galactic sources (i.e. a sharp cutoff above the He II ionizing edge). For simplicity we assume that Γ_{UVB} is constant with redshift, but we note that the resulting proximity zone sizes are insensitive to changes in Γ_{UVB} of a factor of a few, a non-trivial point that we will discuss in more detail in future work.

We compute radiative transfer outputs for quasars of varying luminosity and redshift bracketing the properties of our observed quasar sample in both ionized and neutral scenarios. We assume here that the *overdensity* field does not significantly evolve with redshift from our $z = 6$ output across the redshift range we study, and simply re-scale physical densities by $(1 + z)^3$.

Lyα forest spectra are computed by combining the neutral fraction x_{HI} and gas temperatures from the radiative transfer simulation with the velocity field from the Nyx simulation, summing the absorption from each gas element using the efficient Voigt profile approximation of Tepper-García (2006). Finally, to retrieve R_p from the Lyα forest spectra, we perform a similar analysis as is applied to the real spectroscopic data: we smooth the spectra with a boxcar filter of 20 Å in the ob-

---

[2]Note that the magnitudes for the quasars in our data sample have been calibrated by Bañados et al. (2016) according to the template from Selsing et al. (2016) instead of Lusso et al. (2015). However, due to the similarities between the two templates at observed wavelengths (redwards of Lyα), we expect only a very minor inconsistency in the inferred ionizing luminosity.

Table 11.1 Measurements of the quasar proximity zone sizes, sorted by redshift.

object	z_{em}	M_{1450}	R_p [pMpc]	$R_{p,\mathrm{corr}}$ [pMpc]
PSO J036+03	6.5412	−27.33	3.64 ± 0.13	3.20 ± 0.11
CFHQS J0210-0456	6.4323	−24.53	1.32 ± 0.13	3.47 ± 0.34
SDSS J1148+5251	6.4189	−27.62	4.58 ± 0.13	3.59 ± 0.10
CFHQS J2329-0301	6.417	−25.25	2.45 ± 0.35	4.86 ± 0.70
SDSS J0100+2802	6.3258	−29.14	7.12 ± 0.13	3.09 ± 0.06
SDSS J1030+0524	6.309	−26.99	5.93 ± 0.36	5.95 ± 0.36
SDSS J1623+3112	6.2572	−26.55	5.05 ± 0.14	6.03 ± 0.16
CFHQS J0050+3445	6.253	−26.70	4.09 ± 0.37	4.60 ± 0.41
CFHQS J0227-0605	6.20	−25.28	1.60 ± 1.37	3.15 ± 2.69
PSO J060+24	6.18	−26.95	4.17 ± 1.38	4.26 ± 1.40
CFHQS J2229+1457	6.1517	−24.78	0.45 ± 0.14	1.07 ± 0.33
SDSS J1250+3130	6.15	−26.53	6.59 ± 1.38	7.93 ± 1.66
ULAS J1319+0950	6.1330	−27.05	3.84 ± 0.14	3.77 ± 0.14
SDSS J2315-0023	6.117	−25.66	3.70 ± 1.39	6.26 ± 2.36
SDSS J1602+4228	6.09	−26.94	7.11 ± 1.40	7.28 ± 1.43
SDSS J0303-0019	6.078	−25.56	2.21 ± 0.38	3.88 ± 0.67
SDSS J0842+1218	6.069	−26.91	6.47 ± 0.38	6.71 ± 0.39
SDSS J1630+4012	6.065	−26.19	4.80 ± 0.38	6.59 ± 0.52
CFHQS J1641+3755	6.047	−25.67	3.98 ± 0.38	6.71 ± 0.64
SDSS J2054-0005	6.0391	−26.21	3.17 ± 0.14	4.32 ± 0.19
SDSS J1137+3549	6.03	−27.36	6.98 ± 1.42	6.06 ± 1.23
SDSS J0818+1723	6.02	−27.52	5.89 ± 1.42	4.80 ± 1.16
SDSS J1306+0359	6.016	−26.81	5.39 ± 0.38	5.80 ± 0.41
ULAS J0148+0600	5.98	−27.39	6.03 ± 0.39	5.18 ± 0.33
SDSS J1411+1217	5.904	−26.69	4.60 ± 0.39	5.19 ± 0.44
SDSS J1335+3533	5.9012	−26.67	0.78 ± 0.15	0.89 ± 0.17
SDSS J0840+5624[a]	5.8441	−27.24	0.88 ± 0.15	0.80 ± 0.13
SDSS J0005-0006	5.844	−25.73	2.87 ± 0.40	4.73 ± 0.66
SDSS J0002+2550	5.82	−27.31	5.43 ± 1.49	4.81 ± 1.31
SDSS J0836+0054	5.810	−27.75	5.06 ± 0.40	3.77 ± 0.30
SDSS J0927+2001	5.7722	−26.76	4.68 ± 0.15	5.14 ± 0.16

Notes. The columns show the object's name, its emission redshift, absolute magnitude at 1450 Å in the rest frame, and its proximity zone measurement, as well as the luminosity corrected proximity zone measurement.

[a] Excluded from our analysis due to associated absorbers (see Appendix D).
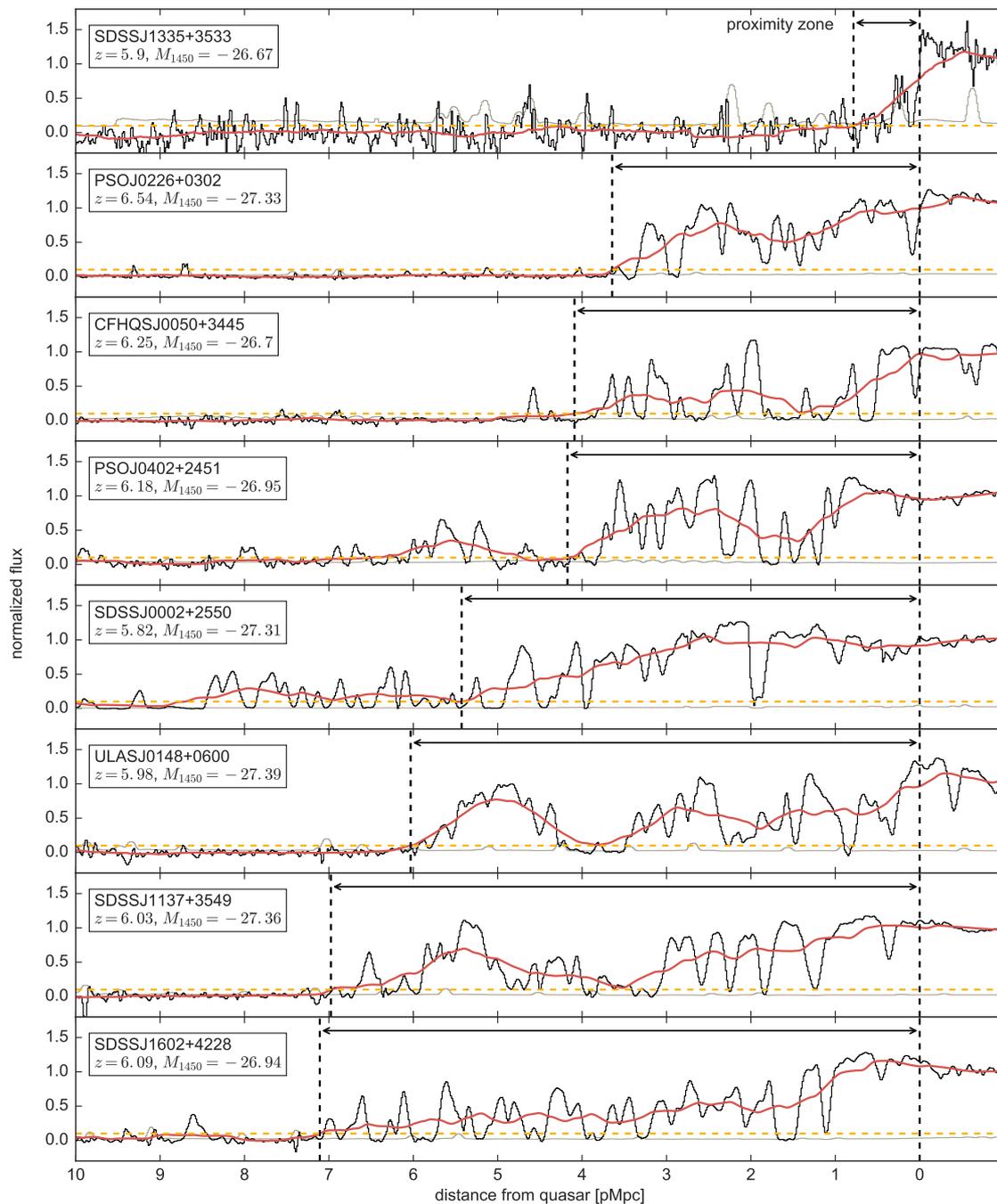
Figure 11.1 Continuum normalized spectra of a subset of bright quasars in our data set with luminosities between $-27.5 \leq M_{1450} \leq -26.5$ showing the transmission profile within 10 pMpc from the quasar. A boxcar smoothing of two pixels has been applied to the spectra (black) and their noise vectors (gray). The red curves show the quasar spectra smoothed to a resolution of 20 Å in observed wavelength. The horizontal yellow dashed lines indicate a flux level of 10%. The vertical black dashed lines show the extent of the quasar proximity zone from the quasar redshift (right) to the first drop of the smoothed flux below the 10% level (left).
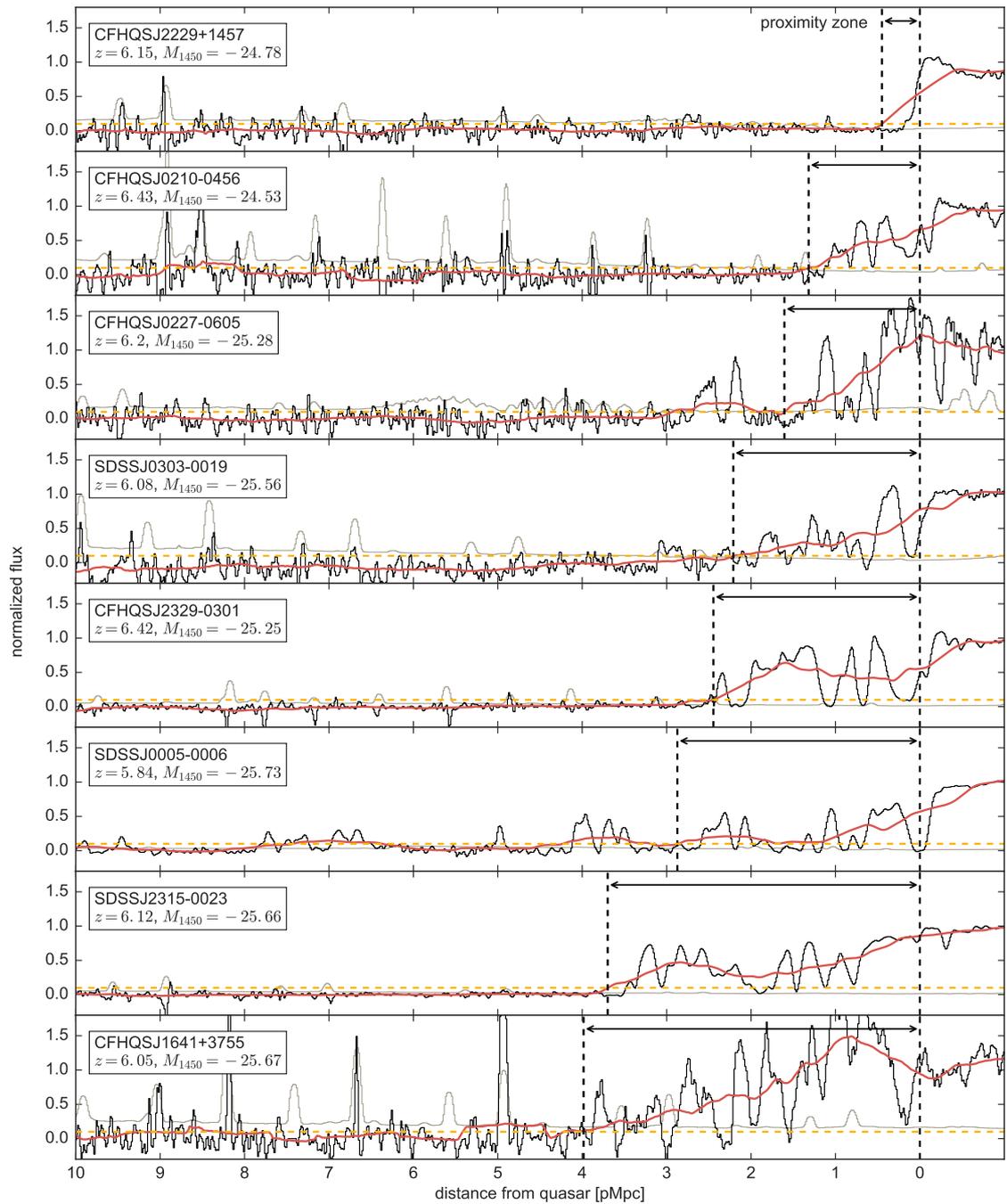
Figure 11.2 Same as Fig. 11.1, for a subset of faint quasars in our data set with luminosities between $-25.8 \le M_{1450} \le -24.5$.

served frame, and then locate where the transmitted flux first drops below 10%. A few skewers through the simulation box giving mock Lyα forest spectra in a highly ionized IGM are shown in Fig. 11.3.
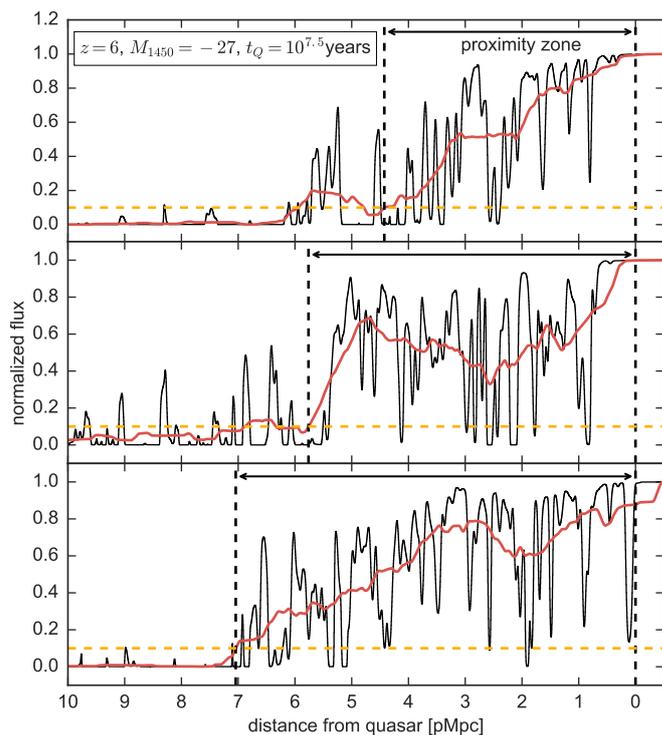
Figure 11.3 Mock spectra from three different skewers through the radiative transfer simulation box for a quasar at $z = 6$, a magnitude of $M_{1450} = -27$, and an age of $t_{\rm Q} = 10^{7.5}$ yr, in a highly ionized IGM. The red curve shows the smoothed spectra and the black dashed lines indicate the extent of the proximity zones.

11.3 The Redshift Evolution of Proximity Zone Sizes

In this section we will present our measurements of the proximity zone sizes for the ensemble of quasar spectra in our data set. Since these quasars cover a wide range of luminosities, we first put them on a common luminosity scale, in order to facilitate a study of the redshift evolution of proximity zone sizes. In the last part of this section we compare our measurements to previous work.

11.3.1 Correcting for the Quasar Luminosity

We expect the size of the proximity zone of a quasar to depend on their luminosity. Proximity zone sizes may also evolve with redshift in a way that tracks the evolution of the ionization state of the IGM driven by the underlying evolution in the UVB. To study the evolution of the proximity zone size with redshift, we have to remove the dependency on quasar luminosity, and normalize our measurements to the same absolute magnitude. The approach that has been adopted in the literature, is to re-scale all quasar luminosities to a common value assuming a particular model for the luminosity dependence, in order to remove the large scatter in proximity zone sizes driven by the wide range of quasar luminosities present in the data set. This then enables one to study the redshift evolution of proximity zone

sizes. However, the scaling of the proximity zone size with luminosity depends on the physical conditions of the ambient IGM.

In this section we study radiative transfer simulations to better understand the expected luminosity scaling, and its dependence on the IGM ionization state. We compute the luminosity scaling in our data set and compare to these simulations. We use these simulations to advocate for an approach to luminosity correct our measurements and search for a redshift evolution.

In general, the luminosity dependence of proximity zone sizes could depend on the ionization state of the surrounding IGM. Specifically Eqn. (6.7) indicates that the size of an ionization front R_{ion} expanding into a neutral IGM scales as $R \propto \dot{N}_\gamma^{1/3}$. On the other hand, using an analytical model Bolton & Haehnelt (2007b) showed that the proximity zone size scales as $R_p \propto \dot{N}_\gamma^{1/2}$, if the IGM is in fact highly ionized by either the quasar itself or the UVB. To better understand this scaling, we turn to our radiative transfer simulations (see § 11.2), which were run for an ensemble of quasars that span the same range of luminosities as our data set. Note that the redshifts of the simulated quasars and the IGM are set to $z = 6$.

The simulated relation between proximity zone size and quasar magnitude for a highly ionized IGM is shown as the gray dashed line in Fig. 11.4, with the shaded region illustrating the 1σ scatter of simulated sizes due to cosmic variance alone (i.e. no measurement error). The relation from the simulations in a highly ionized IGM is reasonably well fit by the power-law

$$R_p \approx 5.57\,\mathrm{pMpc} \times 10^{-0.4 M_{1450}/2.35}. \tag{11.1}$$

Our simulations suggest that the proximity zone sizes in a highly ionized IGM scale as $R_p \propto \dot{N}_\gamma^{1/2.35}$, which lies in between the two theoretically expected relations for a neutral or mostly ionized ambient IGM, shown as the blue dashed-dotted and yellow dotted curves, respectively. Note that the normalization for these two analytic curves has been arbitrarily chosen such that all curves intersect at the same point. The simulated relation does not align exactly with the analytically expected relation in a highly ionized IGM, possibly due to heating effects of He II reionization.

Note however, that we obtain a very similar scaling from our radiative transfer simulations, when assuming a neutral ambient IGM:

$$R_p \approx 5.03\,\mathrm{pMpc} \times 10^{-0.4 M_{1450}/2.45}. \tag{11.2}$$

This is due to the fact that the measured proximity zone sizes R_p do not trace the extent of the ionization front R_{ion} described in Eqn. (6.7) for an assumed quasar

age of $t_Q \sim 10^{7.5}$ yr, but end much earlier within the ionized H II region around the quasar. Because the IGM is indeed highly ionized in this region, one expects a scaling similar to that of a highly ionized IGM. Hence, it is no surprise to find the two relations from the radiative transfer simulations in a highly ionized ambient IGM as well as in an originally neutral ambient IGM to behave similarly. We will come back to this in more detail in § 11.3.3.

Given this new found intuition from radiative transfer simulations, it is now interesting to study the luminosity scaling of proximity zone sizes in our data. To this end, we show the dependence of our measured proximity zone sizes on the quasar's magnitude, which is proportional to the quasar luminosity, in Fig. 11.4. Measurements have been color coded by the emission redshift of the quasar. The errorbars of the measurements reflect the uncertainty on the proximity zone sizes due to the uncertainty of the quasar redshift. We fit a power-law to our measurements and obtain

$$R_p \approx 4.71 \, \mathrm{pMpc} \times 10^{-0.4 \times (M_{1450}+27)/3.42}, \tag{11.3}$$

as the best fit, shown as the red dashed line. The 1σ uncertainty of this fit is given by the shaded region, which was determined by bootstrap re-sampling of our measurements with replacement, and repeating the fit 1000 times. Note that we do not weight our measurements by the redshift errors in this fit. In general, we find reasonable agreement between the scaling relations from the radiative transfer simulations and our data. The data seem to favor a shallower evolution, but this is complicated by the fact that we fit a sample over a large range of redshifts, and also, possibly, that exceptionally small proximity zones (see § 11.4) are pulling down the fit.

So which relation should we adopt to put the measured proximity zone sizes onto a common luminosity scale? Due to the large body of evidence for the IGM being mostly ionized at $z \sim 6$ (e.g. Wyithe & Bolton, 2011; Calverley et al., 2011; Becker et al., 2015b; McGreer et al., 2015) we choose the relation from our radiative transfer simulations assuming a highly ionized IGM, i.e. Eqn. (11.1), to eliminate the dependence on quasar luminosity and normalize our measurements to a common magnitude of $M_{1450} = -27$. However, our results are only marginally dependent on this choice, since the luminosity scaling for an ionized and neutral IGM are in fact very similar, and the main conclusions of this paper are not influenced by this choice. Thus we re-scale our measured proximity zone sizes with

$$R_{p,\mathrm{corr}} \approx R_p \cdot 10^{-0.4(-27-M_{1450})/2.35}. \tag{11.4}$$
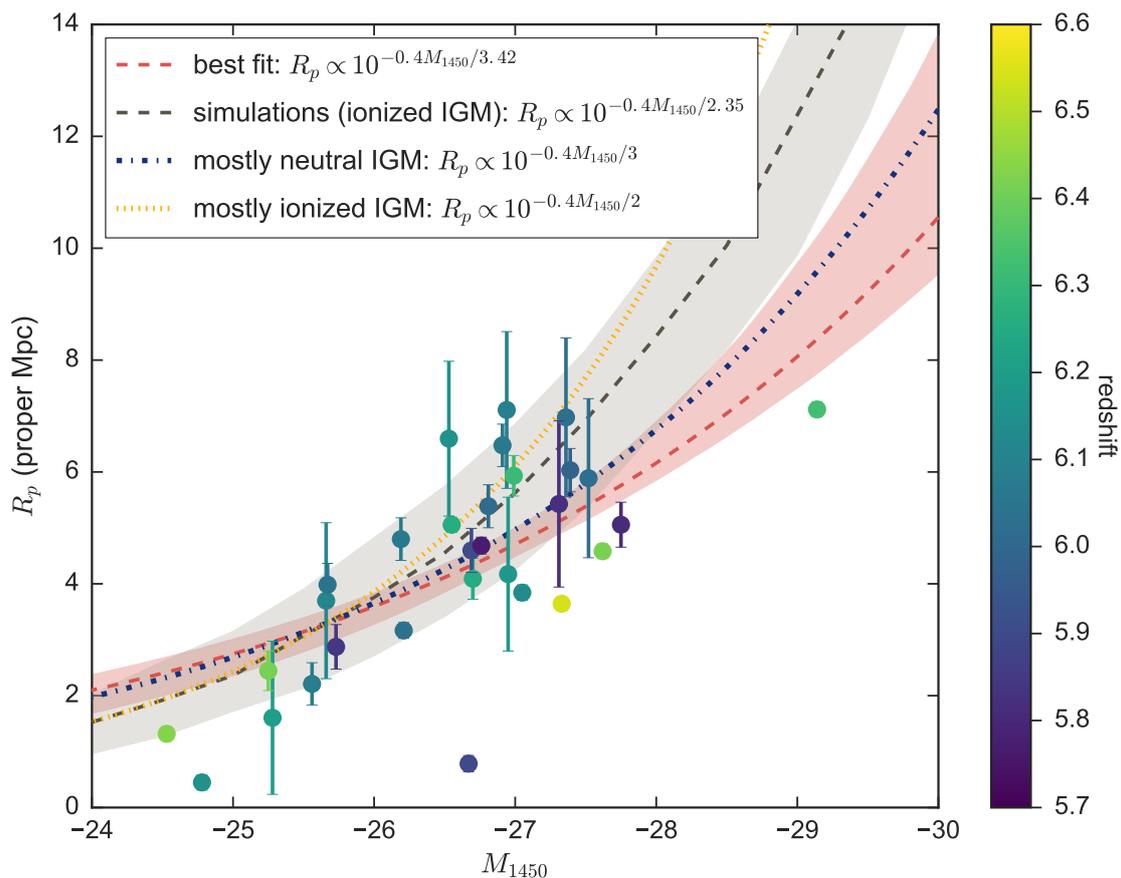
Figure 11.4   Sizes of the proximity zones shown dependent on the quasars' magnitude M_{1450}, color coded by their redshifts. The red dashed line shows the best power-law fit to the measurements with a 1σ-uncertainty level from bootstrapping errors. The gray dashed line shows the expected evolution of the proximity zones from radiative transfer simulations in a highly ionized IGM. The blue dashed-dotted and yellow dotted curves show the theoretical expectations when the IGM surrounding the quasars is neutral or highly ionized, respectively.

Our measurements of the quasar proximity zone sizes are provided in Table 11.1, with the two rightmost columns showing the measured proximity zone R_p in proper Mpc and the corrected proximity zone size $R_{p,\mathrm{corr}}$, which are re-scaled to an absolute magnitude of $M_{1450} = -27$ following Eqn. (11.4).

11.3.2   Measurement of the Redshift Evolution of Proximity Zone Sizes

After correcting our measurements for the quasar luminosity, we can study the evolution of the proximity zones with redshift. In Fig. 11.5 we show this redshift evolution of our luminosity corrected measurements, color coded by M_{1450}, in both
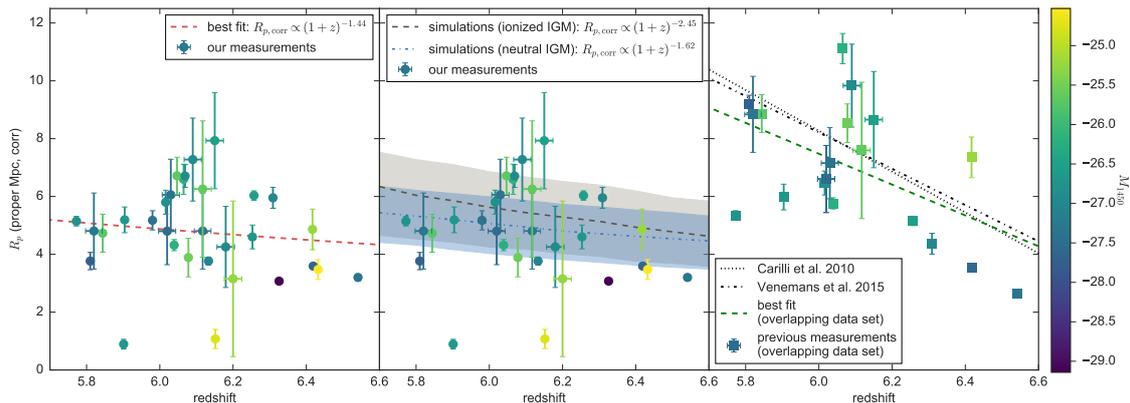
Figure 11.5 Evolution of the luminosity corrected proximity zone sizes with redshift, color coded by the quasars' actual magnitude M_{1450}. The data points showing our measurements for $R_{p,\mathrm{corr}}$ in both the right and middle panel are the same. In the left panel, the red dashed line shows the best power-law fit to the measurements with a 1σ-uncertainty level determined by bootstrapping. In the middle panel the gray dashed line shows the evolution of the proximity zones found in radiative transfer simulations when assuming a highly ionized IGM, whereas the blue dashed line is the result of the radiative transfer simulations assuming a mostly neutral ambient IGM. The shaded areas show the respective 1σ uncertainties of the relations due to cosmic variance. In the right panel the square data points show previous measurements by Carilli et al. (2010) and Venemans et al. (2015) of the quasars in common between our and their data sets, after correcting with updated redshift and magnitude estimates. The green dashed line shows the best fit to these measurements with a 1σ-uncertainty level determined by bootstrapping. The black dotted and dash-dotted curves show linear fits to the measurements of the whole data set from Carilli et al. (2010) and Venemans et al. (2015), respectively.

the left and middle panel. In the left panel, the red dashed line shows the best-fit power-law fit to the redshift evolution

$$R_{p,\mathrm{corr}} \approx 4.87 \, \mathrm{pMpc} \times \left(\frac{1+z}{7} \right)^{-1.44}. \tag{11.5}$$

The shaded region indicates the 1σ uncertainty on the fit determined by bootstrap re-sampling of our measurements with replacement and repeating the fit 1000 times.

In the next section we use our radiative transfer simulations to better understand the redshift evolution of proximity zones, and show that the shallow redshift evolution that we measure is indeed expected. However, the black dashed and dash-dotted lines in the right panel of Fig. 11.5 are linear fits to the measurements of quasar proximity zones from previous analyses of similar data sets by Carilli et al. (2010) and Venemans et al. (2015), which show a much steeper trend. We

compare our measurement to previous work in § 11.3.4, and discuss the sources of this discrepancy.

11.3.3 Understanding the Shallow Redshift Evolution of Quasar Proximity Zone Sizes

We use our radiative transfer simulations to investigate the expected redshift evolution of proximity zone sizes for both a highly ionized ambient IGM and a neutral IGM. Our results are shown in the middle panel of Fig. 11.5 as the gray dashed line (ionized) and blue dash-dotted line (neutral), respectively. In the case of an ionized ambient IGM the corrected quasar proximity zones follow a power-law

$$R_{p,\text{corr}} \approx 5.66 \, \text{pMpc} \times \left(\frac{1+z}{7} \right)^{-2.45} , \tag{11.6}$$

whereas the evolution in the case of a neutral IGM is well fit by

$$R_{p,\text{corr}} \approx 5.06 \, \text{pMpc} \times \left(\frac{1+z}{7} \right)^{-1.62} . \tag{11.7}$$

The shaded regions around the curves show the scatter about these relations due to cosmic variance for quasars with magnitude $M_{1450} = -27$, but they slightly over- and underpredict the scatter for higher and lower luminosity quasars, respectively. The two scenarios of an ionized and a neutral IGM both result in a relatively shallow redshift evolution of proximity zones and are thus consistent with our measured redshift evolution shown in the left panel as the red dashed curve, but inconsistent with the steep evolution found in previous work (right panel).

Why should the redshift evolution of quasar proximity zones be so shallow? The end of the proximity zone around quasars is defined as the location where the smoothed transmitted flux drops below the 10% transmission level and a corresponding limiting optical depth of $\tau_{\text{lim}} = 2.3$ is reached. Bolton & Haehnelt (2007b) determine in Eqn. (8) of their paper the ionization rate Γ_{lim} that is necessary to produce a neutral gas fraction x_{lim} that results in the required limiting optical depth τ_{lim}. Assuming a highly ionized gas in ionization equilibrium and making reasonable assumptions about the gas temperature of the IGM at $z \sim 6$, one obtains an ionization rate of $\Gamma_{\text{lim}} \sim 4 \times 10^{12} \, \text{s}^{-1}$ at the end of the proximity zone R_p, which is an order of magnitude larger than ionization rate of the UVB $\Gamma_{\text{UVB}} \sim 2 \times 10^{-13} \, \text{s}^{-1}$ at this redshift (Wyithe & Bolton, 2011). Thus the total ionization rate at the end of the proximity zone Γ_{lim}, which is the sum of the ionization rate of the background radiation Γ_{UVB} and the ionization rate of the quasar itself Γ_{QSO}, has to be totally dominated by the latter.

As such, in our simulations R_p is essentially independent of Γ_{UVB} and hence fairly insensitive to the neutral fraction x_{HI} of a highly ionized IGM, provided the quasar has been emitting light for longer than the equilibration timescale t_{eq} of the gas, which denotes the timescale on which the gas reaches ionization equilibrium (see § 6.3). The quasar age thus has to be $t_Q \gtrsim t_{\mathrm{eq}} \sim 1/\Gamma_{\mathrm{QSO}}(R_p) \sim 10^5$ yr, in order for the ambient gas to have reached ionization equilibrium.

If the IGM surrounding the quasars is instead very neutral, i.e. $x_{\mathrm{HI}} \sim 0.1 - 1.0$, Eqn. (6.7) for the location of the ionization front indicates that the size of the ionized H II region around the quasar scales as $R_{\mathrm{ion}} \propto (t_Q/x_{\mathrm{HI}})^{1/3}$. This suggests that the proximity zones could be sensitive to the neutral gas fraction (subject to a degeneracy with the quasar age). Indeed for short ages the measured proximity zone R_p will trace the expanding ionization front R_{ion} around the quasar, i.e. $R_p \approx R_{\mathrm{ion}}$, which increases with the quasar age. However, R_p will cease to grow further once it reaches a distance given by the 10% transmission level of the flux, according to the definition of R_p. Thus even as R_{ion} continues to grow with age, the proximity zone size saturates, and will be insensitive to both the quasar age and neutral fraction. This implies that the measured proximity zone size provides a lower limit on the location of the ionization front R_{ion}. The maximum size of the proximity zone $R_{p,\mathrm{max}}$ is given by the distance at which a 10% flux transmission level corresponding to a limiting optical depth of $\tau_{\mathrm{lim}} = 2.3$ is reached. The age of the quasar at this distance is $t_Q > t_Q(R_{p,\mathrm{max}}) \sim 10^6$ yr under the assumption of a homogeneous IGM (see Eqn. (13) in Bolton & Haehnelt, 2007b).

Thus for a highly plausible quasar age of $t_Q \sim 10^{7.5}$ yr (see e.g. Martini, 2004) the proximity zones R_p in a neutral IGM will look the same as for the highly ionized case, because the proximity zones by definition cease to grow once a flux transmission level of 10% is reached, as has been previously pointed out by (Bolton & Haehnelt, 2007b). This demonstrates the insensitivity of R_p to the neutral fraction x_{HI} of the IGM for an appropriate choice of quasar age (see also Khrykin et al., 2016, for an analogous argument in the context of He II proximity zones).

Regarding the redshift evolution of the proximity zone sizes, Eqn. (11) in Bolton & Haehnelt (2007b) predicts a very shallow scaling of $R_p \propto (1+z)^{-2.25}$ due to the density evolution in the universe. The simulations in Fig. 11.5 reveal a similar scaling, i.e. $R_p \propto (1+z)^{-2.45}$ for a neutral ambient IGM and $R_p \propto (1+z)^{-1.62}$ for a highly ionized IGM with a quasar age of $t_Q \sim 10^{7.5}$ yr. The true scaling does not match exactly the analytically expected scaling due to different heating effects.

Hence we conclude that the redshift evolution of R_p is not a very useful probe of the ionization state of the IGM. In a highly ionized IGM due to the definition
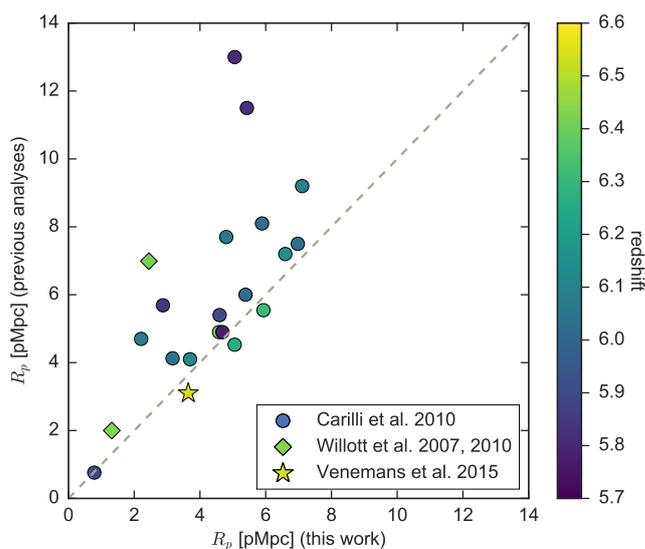
Figure 11.6 Comparison of previous measurements of quasar proximity zones R_p to the analysis presented in this chapter for the overlapping objects in the various data sets. The gray dashed line shows the exact one-to-one relation. In general we recover smaller proximity zones than previous analyses, but mostly consistent within $\lesssim 2$ pMpc.

of proximity zones, the ionization rate at the end of the proximity zone is totally dominated by the ionization rate of the quasar itself and thus essentially independent of the background radiation and the ionization state of the surrounding IGM. In a neutral IGM, R_p ceases to grow once a maximum size $R_{p,\mathrm{max}}$ corresponding to a limiting optical depth τ_{lim} at the 10% flux transmission level is reached. The observed shallow redshift evolution is thus perfectly consistent with our models for both a highly ionized and a neutral IGM for a quasar age of $t_Q > 10^6$ yr, in particular for a fiducial quasar age of $t_Q \sim 10^{7.5}$ yr. None of our simulations reveal the steep evolution that was measured by previous analyses (Carilli et al., 2010; Venemans et al., 2015). Our models do not predict any large proximity zones of $R_p \gtrsim 10$ pMpc that would be consistent with their best-fit curve at $z \approx 5.8$. We will now investigate possible reasons leading to the discrepancy between ours and previous analyses.

11.3.4 Comparison to Previous Analyses

There are several possible reasons causing the differences between ours and previous analyses that result in the different measured redshift evolution of quasar proximity zones: we use a different scaling with luminosity applied to correct all proximity zone measurements and put them on the same luminosity scale, different quasar spectra[3], different approaches to fitting continua and measuring the proximity zone sizes, and our measurements also rely on updated luminosities and

---

[3]Although many of the objects are overlapping between ours and previous data sets, we co-added different exposures from different observing runs and have thus gained higher quality spectra.

redshift measurements. Below we discuss each of these differences and their impact on our results in turn.

To address the issue of different luminosity scalings we facilitate a comparison to previous work by Carilli et al. (2010) and Venemans et al. (2015) by taking their measurements of R_p and correcting them with our luminosity correction ($R_{p,\mathrm{corr}} \propto \dot{N}_\gamma^{1/2.35}$ instead of their $R_{p,\mathrm{corr}} \propto \dot{N}_\gamma^{1/3}$; see Eqn. (11.4)) using updated measurements of M_{1450} from Bañados et al. (2016) for their objects. In their work they chose a linear relation between proximity zone size and redshift, i.e. $R_p \propto z$, and thus we also show a linear fit to their corrected measurements in the right panel of Fig. 11.5 in order to simplify the comparison. Our best fit slope to the measurements of Carilli et al. (2010) is $m \approx -7.07$ (black dotted line) and $m \approx -6.28$ (black dash-dotted line) for measurements by Venemans et al. (2015), which is considerably steeper than the slope $m \approx -1.12$ that we obtain when applying a linear fit to our data. Thus we conclude that the difference in luminosity scaling does not cause the discrepancy between the different analyses.

However, both previous analyses contain objects in their data sample that are not included in our sample. Although the data sets from Carilli et al. (2010) and Venemans et al. (2015) partially overlap with our sample, they also contain distinct objects. In order to verify that differences in the slope of the redshift evolution are not being driven by differences in the quasar samples, we restrict the analysis now to measurements of the 18 quasars that are common to our sample and those of Carilli et al. (2010) and Venemans et al. (2015). For five of these quasars, our analysis uses updated redshift measurements and thus we correct the measurements of R_p from previous work for the difference between the old and the new redshift measurements in order to have consistent redshifts and afterwards use Eqn. (11.4) to obtain $R_{p,\mathrm{corr}}$ from their measurements of R_p. Their corrected measurements of the objects overlapping in both data sets are shown in the right panel of Fig. 11.5 and the best linear fit for the redshift evolution is shown as the green dashed curve. It still shows a relatively steep slope of $m \approx -5.33$ although slightly shallower than before. Thus we conclude that differences in the data sets are not driving the discrepancy in the results.

Note however, that the differences in the resulting redshift evolution of R_p appear to be driven by a handful of objects, where we measure significantly different proximity zone sizes than previous work. In Fig. 11.6 we compare our measurements of the (uncorrected) R_p to those measured by Willott et al. (2007), Willott et al. (2010), Carilli et al. (2010) (who updated the measurements originally performed by Fan et al. (2006c)), and Venemans et al. (2015) for all objects
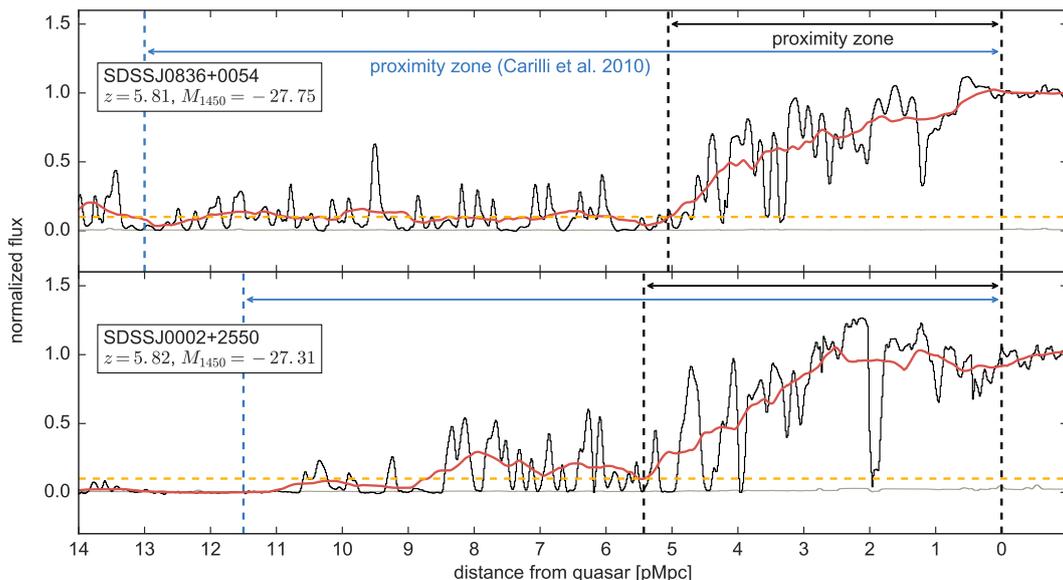
Figure 11.7 Quasar spectra from our data set with the largest discrepancy between our measurements of the proximity zones and the measurements by Carilli et al. (2010).

that are overlapping between our data sample and these previous analyses. We have adopted consistent redshifts for all objects here in order to facilitate a one-to-one comparison that is not driven by changes in redshift, i.e. for the objects for which we have updated redshift measurements, we correct the previously analyzed proximity zones for the difference.

The measurements which lie along the gray dashed line indicate agreement between the different analyses. In general, we measure smaller sizes of the proximity zones than previously obtained, but for most objects the measurements agree within $\Delta R_p \lesssim 2$ pMpc. Despite the fact the fact that for a significant fraction of these overlapping quasars the analyzed data comes from the same instrument (Keck/ESI), small differences in the measurements can be attributed to different data reduction pipelines and in many cases we have co-added data from different runs to get higher S/N spectra and thus the final quasar spectra might differ in quality. However, there are a few outliers in Fig. 11.6, for which our measurements differ significantly from previous work, i.e. $\Delta R_p > 2$ pMpc, which we discuss in further detail.

We measure much smaller proximity zones for three quasars in particular: SDSSJ0836+0054 ($R_{p,\mathrm{Carilli}} \approx 13.0$ pMpc vs. our measurement $R_p = 5.06 \pm 0.40$ pMpc) at $z = 5.810$, SDSSJ0002 + 2550, ($R_{p,\mathrm{Carilli}} \approx 11.5$ pMpc vs. our measurement $R_p = 5.43 \pm 1.49$ pMpc) at $z = 5.82$, and CFHQSJ2329 − 0301, ($R_{p,\mathrm{Willott}} \approx 7.0$ pMpc vs. our measurement $R_p = 2.45 \pm 0.35$ pMpc) at $z = 6.417$. The spectra of the first

two objects, SDSSJ0836+0054 and SDSSJ0002+2550 and our measured proximity zones (black dashed lines) as well as the measurements from Carilli et al. (2010) (blue dashed lines) are shown in Fig. 11.7. Both analyses use the same redshift measurements and thus no differences due to redshift errors should be causing the discrepancy we see here. The spectrum of SDSSJ0836 + 0054 in the upper panel shows transmitted flux throughout the spectrum with the smoothed flux oscillating around the 10% level. Our spectrum nevertheless shows a significant drop below the 10% level at $R_p \approx 5.06$ pMpc. This measured R_p does not change significantly when fitting the continuum model with a different set of PCA components.

For SDSSJ0002 + 2550 (lower panel) the size of the proximity zone is more sensitive to the precise placement of the quasar continuum level. A slightly different continuum normalization (i.e. taking a different set of PCA components) would increase the measurement of the proximity zone from $R_p \approx 5.43$ pMpc to $R_p \approx 8.85$ pMpc, but still falls considerably short of the $R_p = 11.5$ pMpc previous measurement by Carilli et al. (2010). Another reason for the discrepancy in the R_p measurements could be the higher signal-to-noise data in our sample. However, it remains unclear whether the differences in the continuum normalization and the better quality data could cause the whole discrepancy of the two measurements.

The third object that has a significantly different proximity zone measurement is CFHQSJ2329 − 0301, whose proximity zone was determined by Willott et al. (2007) to be $R_p \approx 6.3$ pMpc (after updating their measurements with a new redshift measurement it is now $R_p \approx 7.0$ pMpc). However, most of this discrepancy between their measurement and ours can be contributed to the fact that for this particular object, Willott et al. (2007) decided after inspecting the spectrum by eye that they take the *second* drop of the smoothed flux below 10% level as the end of the proximity zone instead of the *first* drop according to the standard definition of proximity zones (Fan et al., 2006c). They determine the first drop in flux below the 10% transmission level to be at $R_p \approx 3.7$ pMpc (with an updated redshift measurements this would be $R_p \approx 4.4$ pMpc), which would be much more consistent with our measurement of $R_p = 2.45 \pm 0.35$ pMpc.

Another reason for the discrepancy in the redshift evolution of the proximity zone measurements is caused by the exclusion of the quasar SDSSJ1335 + 3533 at $z = 5.9012$ from previous analyses. For this particular object, which has a very small proximity zone of $R_p = 0.78 \pm 0.15$ pMpc, our measured size agrees with the one measured by Carilli et al. (2010). However, this object is a weak emission line quasar and has been somewhat arbitrarily excluded from previous analyses due to this fact (see Carilli et al., 2010; Wyithe & Bolton, 2011). The reason given in

the aforementioned papers is the "fundamentally different nature" of such objects. However Diamond-Stanic et al. (2009) showed that weak emission line quasars do not show significantly different UV continuum slopes apart from the emission lines and thus do not differ in their physical properties. The original reason for Fan et al. (2006c) to exclude this object from their analysis was that the only redshift measurement they had at that time came from the onset of strong Lyα absorption. Given that we now have an accurate redshift measurement, we include SDSS J1335 + 3533 into our analysis, which causes the slope of the redshift evolution of the proximity zones to become shallower due to its small zone. We will further discuss the implications of small proximity zones in § 11.4.

Further differences between the different analyses can be attributed to the different continuum fitting methods. Previous analyses applied a power-law fit to the quasar continuum redwards of the Lyα emission line and fitted the Lyα and N V lines with Gaussian curves, whereas we chose to model the continuum of each quasar with a PCA (Pâris et al., 2011). Although differences in the continuum estimation can change the size of individual proximity zones, we do not expect them to alter the distribution of proximity zone sizes, since the scatter on R_p resulting from continuum uncertainties is much smaller than the intrinsic scatter due to density fluctuations (Kramer & Haiman, 2009).

To conclude, it is not completely clear to us, how other authors obtained some very large proximity zones, particularly for a few objects at $z < 6$ that are the main drivers for the steep redshift evolution. Some of the differences between ours and previous analyses and the resulting shallower redshift evolution of the proximity zone sizes can be attributed to a variety of reasons: first, we stick to a rigorous definition of the proximity zone sizes and continue to take the first drop of the smoothed flux below the 10% level, since we adopt the same rigorous treatment for the mock spectra from the radiative transfer simulations. Second, we analyze higher quality data, which can help to determine the drop in flux below the 10% level easier. Third, we include all quasars of the ensemble in our analysis and do not exclude the weak emission line quasar in our sample. Finally, we analyze a larger sample of quasar spectra than previous analyses. Differences in the continuum estimation, updated redshift and magnitude measurements, and the different correction for the quasar luminosity do not have a significant influence on the discrepancy.
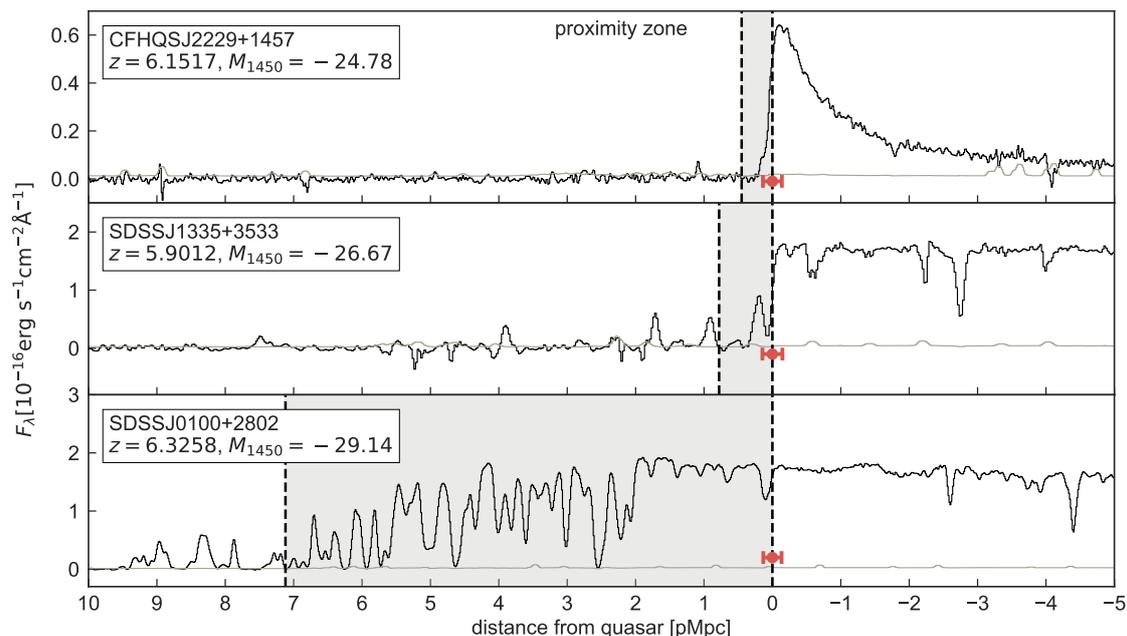
Figure 11.8 Spectra of the three quasars CFHQSJ2229 + 1457 (upper panel), SDSSJ1335 + 3533 (middle panel), and SDSSJ0100 + 2802 (lower panel), exhibiting exceptionally small proximity zones, shown as the grey shaded region. The red errorbars demonstrate the uncertainty on the proximity zone measurements arising from the redshift uncertainty.

11.4 Quasars with Exceptionally Small Proximity Zones

Several of the quasars we studied have particularly small proximity zones, as can be seen from Fig. 11.5. The proximity zones of two objects, i.e. CFHQSJ2229 + 1457 and SDSSJ1335 + 3533, are $R_p < 1$ pMpc and $R_{p,\mathrm{corr}} \lesssim 1$ pMpc. Additionally, given its extreme brightness, the proximity zone of SDSSJ0100 + 2802, $R_p = 7.12 \pm 0.13$ pMpc and $R_{p,\mathrm{corr}} = 3.09 \pm 0.06$ pMpc, is also exceptionally small. In this section, we discuss the properties of these objects and possible explanations for their small proximity zones.

11.4.1 Notes on Individual Objects

CFHQS J2229+1457

This relatively faint quasar, $M_{1450} = -24.78$ (Bañados et al., 2016), was discovered in the Canada-France High-z Quasar Survey (CFHQS) and published by Willott et al. (2009). The measurements of the proximity zone are $R_p = 0.45 \pm 0.14$ pMpc and, when normalized to a magnitude of $M_{1450} = -27$, $R_{p,\mathrm{corr}} = 1.07 \pm 0.33$ pMpc.

The quasar has a precise redshift measurement of $z = 6.1517$ from [C II] emission from the host galaxy (Willott et al., 2015) and hence a very small redshift uncertainty $\Delta v = 100$ km/s resulting in an uncertainty of the proximity measurement of $\Delta R_p \approx 0.14$ pMpc. The measurements of the proximity zone are independent of the choice of PCS used to model the quasar continuum, i.e. continuum uncertainties do not influence our measurements. The top panel of Fig. 11.2 shows the continuum normalized spectrum and the proximity zone of this object. In the upper panel of Fig. 11.8 we show the whole spectrum, revealing no BAL features.

SDSS J1335+3533

This quasar that has been discovered in SDSS by Fan et al. (2004) has a precise redshift measurement of $z = 5.9012$ from CO $(6 - 5)$ emission from the host galaxy (Wang et al., 2010), and is fairly bright with a magnitude of $M_{1450} = -26.67$ (Bañados et al., 2016). The measured proximity zone size for this object is $R_p = 0.78 \pm 0.15$ pMpc and the luminosity corrected size of the proximity zone is $R_{p,\mathrm{corr}} = 0.89 \pm 0.17$ pMpc. The redshift uncertainty of $\Delta v = 100$ km/s results in an uncertainty of the proximity zone measurement of $\Delta R_p \approx 0.15$ pMpc. These measurements of the proximity zone are again independent on continuum modeling uncertainties. The top panel of Fig. 11.1 shows the continuum normalized spectrum and the proximity zone of this object. The whole spectrum of the quasar shown in the middle panel of Fig. 11.8 is completely devoid of broad emission lines, as previously noticed from its discovery spectrum by Fan et al. (2004). The spectrum does not show any BAL features.

SDSS J0100+2802

This object is the brightest high redshift quasar known so far (Wu et al., 2015) with an absolute magnitude of $M_{1450} = -29.14$ (Bañados et al., 2016). Its redshift, $z = 6.3258$, has been measured precisely by the detection of the [C II] line (Wang et al., 2016). Given its extreme brightness the quasar exhibits a very small proximity zone of $R_p = 7.12 \pm 0.13$ pMpc ($R_{p,\mathrm{corr}} = 3.09 \pm 0.06$ pMpc, when normalized to a luminosity of $M_{1450} = -27$) compared to a proximity zone size of $R_p = 12.0 \pm 2.0$ pMpc, that one would expect to see for an object this bright. The redshift uncertainty $\Delta v = 100$ km/s results in an uncertainty of the proximity zone measurement of $\Delta R_p \approx 0.13$ pMpc. Continuum uncertainties do not play any significant role for these measurements. The lower panel of Fig. 11.8 shows the spectrum of this object, which does not show any BAL features.
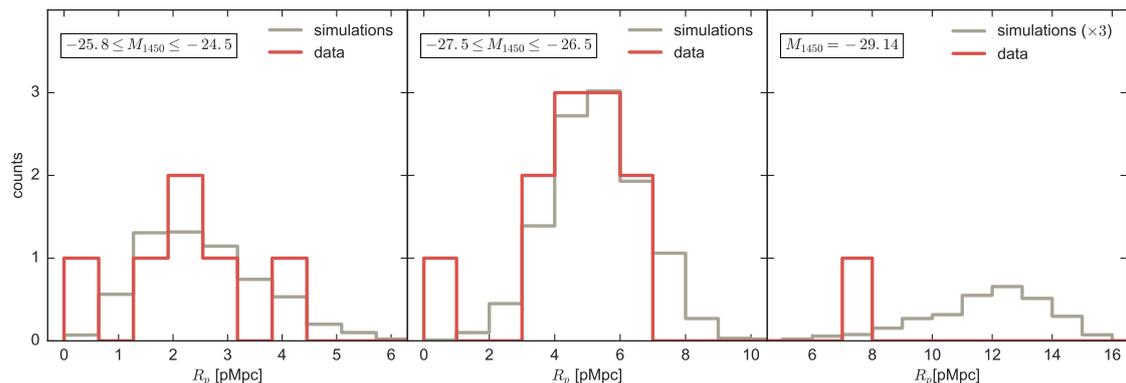
Figure 11.9 Distribution of sizes of quasar proximity zones for faint quasars with luminosities between $-25.8 \leq M_{1450} \leq -24.5$ (left panel), brighter quasars with luminosities between $-27.5 \leq M_{1450} \leq -26.5$ (middle panel) and for very bright quasars with luminosity $M_{1450} \approx -29.14$ (right panel). The red histogram shows the distribution of the measured proximity zone sizes of quasars in our data sample within the given luminosity range that have redshift measurements from Mg II, [C II] or CO lines. The gray histogram shows the expected distribution for quasars with the same redshifts and luminosity properties as in the data sample from the radiative transfer simulations. In the right panel the simulated distribution is multiplied by a factor of three for better visibility. Our data sample includes six fainter quasars with these requirements, eleven brighter ones, and one very bright one.

11.4.2 How Common are Small Proximity Zones?

In order to quantify the probability of finding objects with such small proximity zones, we compare in Fig. 11.9 our measured proximity zones R_p from a subset of quasar spectra from our data sample to proximity zones from our radiative transfer simulations. We show the distribution of quasar proximity zone sizes for three different luminosity ranges, quasars with magnitudes $-25.8 \leq M_{1450} \leq -24.5$ are shown in the left panel, brighter objects with magnitudes $-27.5 \leq M_{1450} \leq -26.5$ are shown on the middle panel, and the right panel shows the distribution of proximity zones for very bright quasars with $M_{1450} = -29.14$. The magnitude intervals were chosen, such that each bin has a width of ~ 1 dex and includes one of the small proximity zone objects. This results in six quasars in the faintest magnitude range, eleven quasars in the brighter range of magnitudes and only one object at $M_{1450} \approx -29$ (Wu et al., 2015). Because of this split into magnitude intervals, we are comparing R_p instead of $R_{p,\mathrm{corr}}$, and model the ranges of quasar luminosities in the simulations. Note that the redshift range covered by the quasars in each bin is very broad, i.e. $5.844 \leq z \leq 6.4323$ for the faintest magnitude bin, $5.7722 \leq z \leq 6.5412$ for the brighter magnitude range, and the only one bright quasar has a redshift

107

$z \approx 6.3258$. We forward model the redshift evolution of the quasars in the radiative transfer simulations, such that the broad range of luminosities in each bin does not influence our results.

The red histograms show the distribution of measured proximity zones from our data set. In this figure we only consider quasars that have their redshifts determined by the detection of a Mg II, [C II] , or CO line, i.e. with small redshift uncertainties. These histograms have been normalized to the number of objects available in our data set within the given magnitude interval. The gray histograms show the distribution of proximity zone sizes expected from our radiative transfer runs. We simulate multiple realizations of each quasar in our sample matching the redshift, luminosity, and redshift uncertainties (which we add as a Gaussian uncertainty to the quasar position in our mock spectra; see Fig. 11.3). For the two leftmost panels of Fig. 11.9 we simulate 100 skewers for each quasar in our data sample. Thus the gray histograms show the proximity zones of 600 skewers in the left panel and 1100 skewers in the middle panel. For the single very bright quasar in the right panel of Fig. 11.9 we simulate 900 realizations to improve the statistics.

We do not model continuum uncertainties in the simulations, which we expect to have negligible influence on the distribution of proximity zone sizes (see discussion in § 11.3.4, Kramer & Haiman, 2009). For the radiative transfer runs we assume a highly ionized IGM, which is consistent with empirical constraints on the neutral gas fraction at $z \sim 6$ (Calverley et al., 2011; Wyithe & Bolton, 2011; Becker et al., 2015b; McGreer et al., 2015). Following the arguments of § 11.3.3 we do not expect any qualitative changes when assuming a mostly neutral IGM.

In both the faint and the brighter case (left and middle panel) the bulk of the distributions of measured and simulated proximity zones agrees very well. Note that our simulations do not reproduce the very large proximity zones of $R_p \sim 10 - 13$ pMpc that Carilli et al. (2010) measured previously. However, the simulations also do a poor job of reproducing the frequency of very small proximity zones that we find. For the fainter quasars (left panel) our simulations indicate that the probability of finding a quasar with $R_p \approx 0.45$ pMpc, which is our measurement for CFHQSJ2229 + 1457, in a sample of six quasar spectra is $\approx 3\%$. However, the significance of this small proximity zone is limited by the definition of the proximity zone. The smoothed flux in the spectrum of CFHQSJ2229 + 1457 remains below the $< 10\%$ level within a distance of 10 pMpc to the quasar, whereas a by eye inspection of the simulated small proximity zones reveals that in most spectra the flux increases above 10% again just outside of their proximity zone. If we would adopt a different definition of the proximity zone that would also be sensitive to the

length of the GP trough outside of the proximity zone, this object would be an even greater outlier.

In the middle panel showing quasars with magnitudes $-27.5 \leq M_{1450} \leq -26.5$ we have one object, SDSS J1335 + 3533, with a measured proximity zone $R_p \approx 0.78$ pMpc. In the sample of 1100 simulated quasars with similar redshifts and magnitudes only one of them has a proximity zone size that small. Thus the probability of finding such an object in a sample of eleven quasars is $\approx 1\%$. In the right panel we have only one object in our data set, SDSS J0100 + 2802. Its proximity zone is relatively small given its extreme brightness. We estimate the probability of finding a quasar with a proximity zone of $R_p \lesssim 7.12$ pMpc to be $\approx 3\%$.

In summary, in all three respective magnitude ranges, the occurrence of small proximity zones is much higher in our data sample than the simulations predict. We will now investigate several possible scenarios that could explain the exceptionally small proximity zones of the three aforementioned objects.

11.4.3   Possible Explanations for Small Proximity Zones

We address three scenarios that could possibly explain the exceptionally small proximity zones of the three quasars mentioned in the previous subsection: The proximity zones could be prematurely truncated due to associated dense absorbers, such as DLAs or Lyman Limit systems (LLSs), patches of remaining neutral hydrogen within the IGM could truncate the proximity zones, or the quasars could be very young.

Truncation Due to Damped Lyα Systems or Lyman Limit Systems

One possible explanation for the exceptionally small proximity zone sizes could be the truncation of the zones due to strong absorbers, such as DLAs or LLSs, which could either be intervening or physically associated with the quasar environment. The presence of such self-shielding absorbers could prematurely truncate the proximity zone at small radii by blocking the quasar's ionizing flux. These optically thick absorption line systems are not correctly modeled by our radiative transfer simulations (see § 11.2). We model only the gas densities in the IGM, but do not include such dense gas patches that could be coming from galaxy like overdensities resulting in these absorbers. Thus, proximity zones truncated prematurely by dense absorption systems are not reproduced in our simulations and need to be identified and eliminated from any comparison between the data and the simulations. To this end, we search for signatures of strong absorption line systems
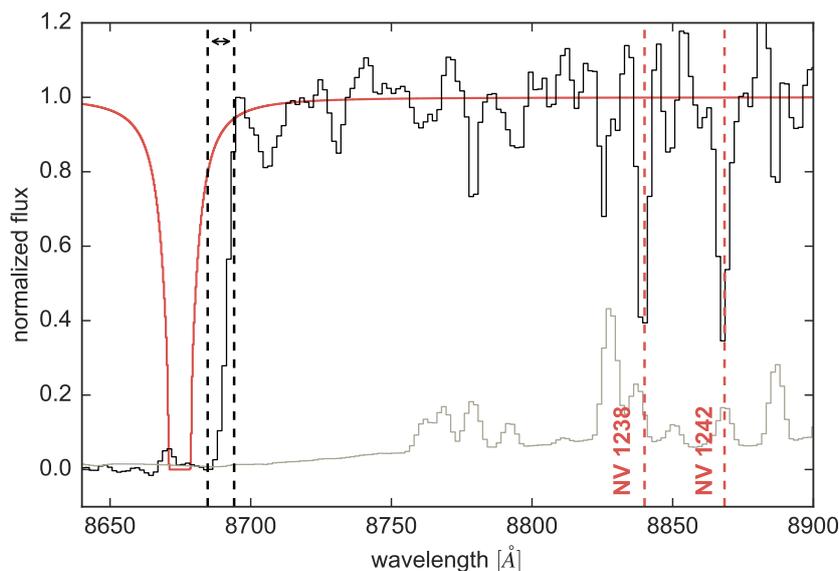
Figure 11.10 Excerpt of the spectrum of CFHQS J2229 + 1457 taken with Low Resolution Imaging Spectrometer (LRIS). We find an N V absorption line system (red dashed lines) associated with an absorber at $z \approx 6.136$ (red curve) with a column density $N_{HI} = 10^{19} \mathrm{cm}^{-2}$ and Doppler parameter $b = 40$ km/s. A boxcar smoothing of two pixels has been applied to both the spectrum (black curve) and its noise vector (gray curve). The vertical black dashed lines indicate the extent of the proximity zone.

in the quasar spectra near the end of their proximity zones by visually inspecting the spectra, and searching for evidence of damping wings which would indicate the presence of a strong absorber, as well as associated ionic metal-line transitions.

In the spectrum of SDSS J1335 + 3533, the only strong metal line absorber we could find was a Mg II absorber at $z \approx 2.10$. However, we have to note that the signal-to-noise ratio of the ESI spectrum is insufficient to identify very weak metal absorption features. We will address this particular quasar again in the following chapter and conduct an extensive search for associated absorption systems with better quality data (see § 12.3).

Searching the ESI spectrum of CFHQS J2229 + 1457 for metal absorption lines associated with nearby dense absorption systems did not reveal any absorbers due to the very low signal-to-noise data (S/N ≈ 2 per pixel). To facilitate the search for metal lines, we obtained a higher quality spectrum with the LRIS at the Keck I telescope with an exposure time of 3600 s. The observations were conducted on September 16th, 2016 using the 600/7500 grating and a slit width of 1", resulting in a resolution of $R \approx 1800$, which is lower than the ESI spectrum ($R \sim 4000$), but it has a higher signal-to-noise ratio (S/N ≈ 7 per pixel). Note that we only use this LRIS spectrum for follow-up analysis, i.e. searching for metal absorption lines. For
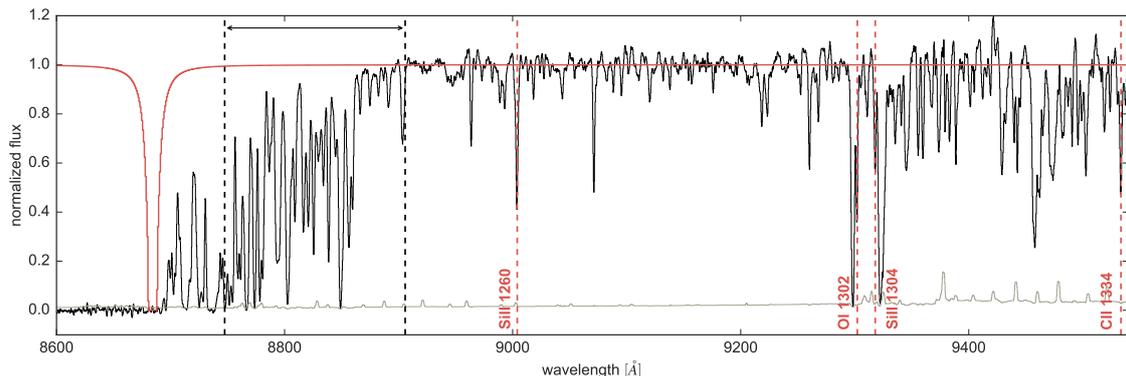
Figure 11.11 Excerpt of the spectrum SDSS J0100 + 2802. We find a close low ionization absorption system at $z \approx 6.144$ (red curve) with column density $N_{\mathrm{HI}} = 10^{19} \mathrm{cm}^{-2}$ and Doppler parameter $b = 40$ km/s and associated absorption lines (red dashed lines). The continuum normalized spectrum (black curve) and its noise vector (gray curve) are smoothed with a three pixel boxcar function. The vertical black dashed lines indicate the extent of the proximity zone. Note that in this figure we show the quasar spectrum normalized with a hand-fitted continuum.

the main analysis of the proximity zone of this quasar we use the ESI spectrum, in order to be consistent with the remaining data sample.

An excerpt of this spectrum is shown in Fig. 11.10 and reveals an N V doublet associated with an absorber at $z \approx 6.136$, whose H I absorption Voigt profile is shown as the red curve when assuming a column density of $N_{\mathrm{HI}} = 10^{19}$ cm^{-2} and a Doppler parameter of $b = 40$ km/s. The absorber is located at a distance of 0.91 pMpc from the quasar and thus it is highly unlikely that this absorber influences the size of the proximity zone significantly, because the proximity zone ends at a higher redshift of $z \approx 6.144$ at a distance $R_p \approx 0.45$ pMpc. Additionally, at the location of the absorber in the spectrum, there is a transmitted flux spike in the Lyα forest partly visible (at $\lambda \approx 8670$ Å), which indicates that the absorber cannot be saturated, implying $N_{\mathrm{HI}} \lesssim 10^{14}$ cm^{-2}, i.e. its column density would need to be much less than the here assumed $N_{\mathrm{HI}} = 10^{19}$ cm^{-2} in order not to violate the spectrum, and is thus unlikely to truncate the proximity zone. We also do not find evidence for any low-ion absorption lines, such as Si II at $\lambda_{\mathrm{rest}} = 1260.42$ Å, that should be present in DLAs.

In the spectrum of SDSS J0100 + 2802 we find a low ionization absorption system close to the object at $z \approx 6.144$ (see Fig. 11.11), which has previously been identified by Wu et al. (2015). However, this absorber is at such large line-of-sight distance $R \approx 9.94$ pMpc from the quasar that it cannot truncate its proximity zone, that already ends at a distance of $R_p = 7.12 \pm 0.13$ pMpc. We show the H I Voigt absorption profile (red curve) assuming an absorber with $N_{\mathrm{HI}} = 10^{19}$ cm^{-2} and

111

$b = 40$ km/s, and its associated lines (Si II at $\lambda_{\rm rest} = 1260.42$ Å and $\lambda_{\rm rest} = 1304.37$ Å, O I at $\lambda_{\rm rest} = 1302.16$ Å, [C II] at $\lambda_{\rm rest} = 1334.53$ Å). Note that an absorption system with a higher column density of $N_{\rm HI} \sim 10^{20.3}$ cm^{-2} characteristic of DLAs, would be too broad and would be inconsistent with the presence of flux transmission that are observed in the Lyβ forest at this redshift.

In light of the possibility that proximate absorbers could truncate the proximity zones, it is interesting to estimate the number of those proximate absorbers that one would expect to find around quasars. Therefore, we start with an estimate of intervening absorbers along the line-of-sight to quasars by Songaila & Cowie (2010), who have estimated the number density dN/dz of intervening LLSs in Fig. 5 and Table 3 of their paper. They determine the best power-law fit to their measurements with a maximum likelihood analysis to be

$$\frac{{\rm d}N}{{\rm d}z} \approx 2.84 \left(\frac{1+z}{4.5} \right)^{2.04} , \tag{11.8}$$

which results in an estimate of dN/d$z \approx 6.99$ at $z = 6$.

On the one hand the abundance of LLSs could be enhanced in the proximity of quasars because of the overdense quasar environment (Hennawi et al., 2006; Prochaska et al., 2008, 2013). However, the intense radiation from the quasar will also ionize the dense gas in its surroundings making it less likely to self-shield, possibly lowering the abundance of so-called proximate LLSs (Hennawi & Prochaska, 2007; Prochaska et al., 2008). Prochaska et al. (2010) found empirically that at $z \sim 4$ the number density of proximate LLSs (occurrence within $\Delta v \leq 3000$ km s^{-1} of the quasar emission redshift) roughly equals the number density of intervening LLSs (see Fig. 15 of Prochaska et al., 2010). Assuming the same approximate equality holds at $z \sim 6$, we estimate the probability p of finding a proximate LLS within 1 pMpc to a quasar at $z = 6$ to be $p \approx n(z)\Delta z \approx 0.11$. Given the size of our data set, which consists out of 30 quasars, we would expect to find a proximate LLS in ~ 3 spectra.

Where are the ~ 3 proximate LLSs that we expect to find in our data sample? We found only one in the spectrum of the quasar SDSSJ0840 + 5624, which we excluded from our analysis for this reason (see Appendix D). The absence of other proximate LLSs detected in our data sample could be attributed to a few different reasons: the estimate of proximate LLSs by Prochaska et al. (2010) assumes a $\Delta v \leq 3000$ km s^{-1} window, which corresponds to a distance of ≈ 4.3 pMpc to a quasar at $z = 6$. We are just interested in the innermost 1 pMpc to the quasar, where the quasar's radiation is stronger and could thus have photoionized all possible LLSs (Hennawi & Prochaska, 2007). A second possible reason for the lack of detections of
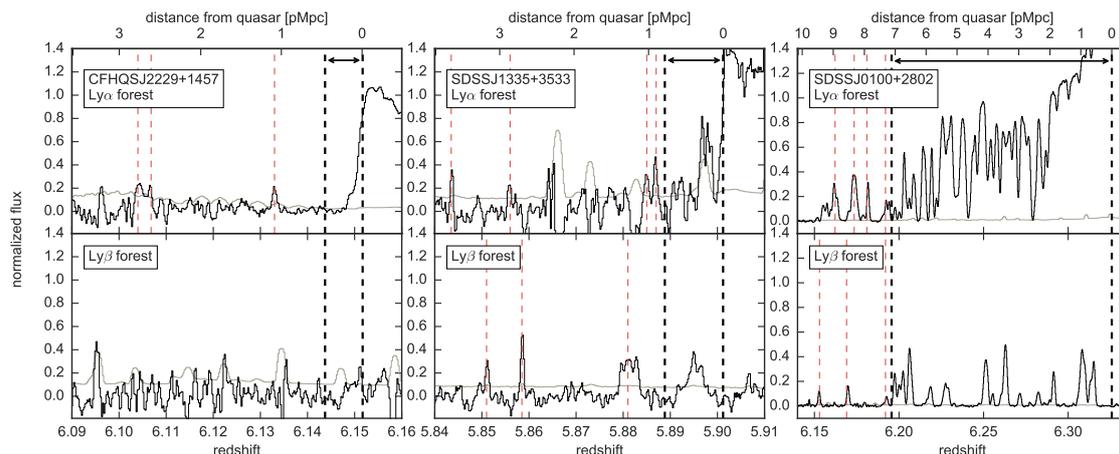
Figure 11.12 Excerpts of the Lyα and the Lyβ forests of the three spectra of CFHQSJ2229 + 1457 (left panels), SDSSJ1335 + 3533 (middle panels) and SDSSJ0100 + 2802 (right panels). The panels show the Lyα and the Lyβ forest (upper and lower panels, respectively) of the spectra at the same redshift and distance close to the quasars. A boxcar smoothing of three pixels has been applied to both the spectra (black curves) and noise vectors (gray curves). The extents of the proximity zones are indicated by the vertical black dashed lines. Red dashed lines indicate transmission spikes outside of the respective proximity zones.

more proximate LLSs in our sample, could be due to the fact that our data sample at $z \sim 6$ is brighter than the one analyzed by Prochaska et al. (2010) at $z \sim 3.5$ and thus the higher radiation could have photoionized more LLSs. Hence, it might be not surprising, that we are only detecting one proximate LLS in our data sample instead of the expected ~ 3.

Although we cannot rule out the presence of absorbers with weak metal lines or rare metal-free absorbers (Fumagalli et al., 2011; Simcoe et al., 2012; Cooper et al., 2015; Crighton et al., 2016; Cooke et al., 2017) in all objects, which would require higher quality data, we do not see evidence for proximate LLSs in the three spectra showing small proximity zones, particularly in the spectrum of the very bright quasar SDSSJ0100 + 2802, where we would most likely be able to detect it.

Neutral Islands of Gas in a Patchily Reionized IGM

Another possible explanation for the existence of very small proximity zones could be the truncation of the proximity zones due to islands of very neutral gas in an otherwise highly ionized IGM, which can occur in patchy reionization models (e.g. Lidz et al., 2007; Mesinger, 2010). This scenario implies that reionization is not yet complete at $z \sim 6$ and these quasars lie in (presumably rare) neutral environments that could absorb all incoming flux and thus truncate the proximity zones. Opposite

to the previously described scenario where LLSs, i.e. dense gas on galactic scales, truncate the proximity zones, we assume here neutral patches of ~ 10 pMpc scales that are associated with reionization and do not have any low redshift counter parts. This explanation requires the presence of large patches of highly neutral gas just outside the proximity zone, which would also absorb flux in the Lyman-series (Lyα, Lyβ, etc.) forests. In other words, in a patchy reionization scenarios one expects to see long GP troughs that are line black in both Lyα and Lyβ, beginning right at the edge of the proximity zone.

In Fig. 11.12 we show continuum normalized spectra of the Lyα and the Lyβ forest (top and bottom panels, respectively) near the three quasars in question. The black dashed lines in all panels indicate the extent of the quasar proximity zone. The left panels show the Lyα and the Lyβ forest close to the quasar CFHQSJ2229+ 1457. While we do not detect any transmitted flux in the Lyβ forest, we do see a few transmission spikes in the Lyα forest at $R \approx 1.1$ pMpc and $R \approx 2.8$ pMpc, indicated by the red dashed lines. The transmission in the Lyβ forest at the same distances must have been absorbed due to foreground Lyα absorption at $z \approx 5.02$ and $z \approx 4.99$.

The middle panels show the Lyα and Lyβ forest of SDSSJ1335+3533, where we see transmitted flux in the Lyα forest just outside of the proximity zone, i.e. at $R \approx 0.9$ pMpc and $R \approx 1.0$ pMpc, and a bit further away at $R \approx 2.9$ pMpc and $R \approx 3.6$ pMpc. In the Lyβ forest we can see clear transmission spikes just outside of the proximity zone at $R \approx 1.4$ pMpc, $R \approx 2.8$ pMpc and $R \approx 3.3$ pMpc (see § 12.3 for a more detailed analysis of this quasar spectrum).

The right panels show excerpts of the Lyα and Lyβ forest for SDSSJ0100+ 2802. One can clearly see several transmitted flux spikes in the Lyα forest right outside the proximity zone. Also the Lyβ forest shows a few flux spikes close to the proximity zone, although less prominent ones.

These detections of transmitted flux just outside the proximity zones in the Lyα or Lyβ forest suggest that if there are patches of neutral gas present close to the quasar, they need to be $\lesssim 1$ pMpc in size, corresponding to the size of the regions that are both line black in the Lyα and Lyβ forest. Although we cannot completely rule out a patchy reionization scenario with such small islands of neutral gas, our constraints on their sizes already suggest that they would be difficult to distinguish from metal poor LLSs. Deeper spectra covering their proximity zones further down the Lyman series could reveal additional Lyβ or Lyγ transmission spikes that would further constrain this scenario, as well as enabling more sensitive searches for metal lines to rule out an LLS scenario (see § 12.3).
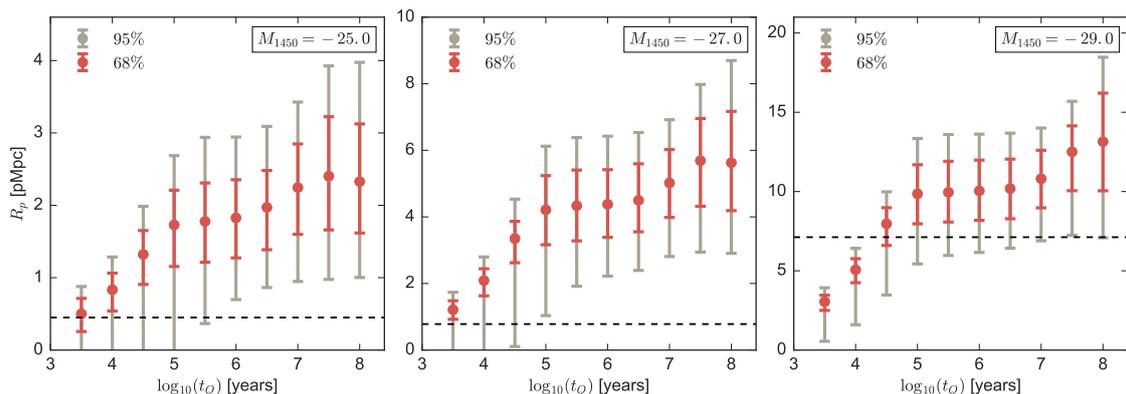
Figure 11.13 Distributions of proximity zone sizes shown as a function of the average quasar age t_Q for quasars with magnitudes of $M_{1450} = -25.0$ (left panel), magnitudes of $M_{1450} = -27.0$ (middle panel) and magnitudes of $M_{1450} = -29.0$ (right panel). The proximity zone sizes of CFHQSJ2229+1457, SDSSJ1335+3533 and SDSSJ0100+2802 are shown as the dashed black line in the left, middle and right panel, respectively.

Young Quasar Age

Finally, another possibility to explain small proximity zones is that the quasars may have turned on only very recently. When a quasar turns on the IGM requires a finite amount of time to adjust to the new higher photoionization rate. We will focus the discussion in this section on a highly ionized surrounding IGM, consistent with optical depth measurements at $z \sim 6$ (e.g. Wyithe & Bolton, 2011; Calverley et al., 2011; Becker et al., 2015b; McGreer et al., 2015; Bosman et al., 2018; Eilers et al., 2018a), although we note similar arguments apply in the neutral case (see Khrykin et al., 2016, for further details).

In § 6.3 we have argued that a quasar has to have been shining for at least $t_Q \gtrsim t_{eq}$, in order for the surrounding IGM to have reached ionization equilibrium with the quasar and result in the maximum proximity zone size. Ionizing equilibrium is achieved more rapidly close the quasar due to the stronger radiation field ($\Gamma_{QSO} \propto R^{-2}$). In order to obtain an optical depth of $\tau_{lim} = 2.3$ at $z \sim 6$, which corresponds to a transmitted flux level of 10% in the Lyα forest, a value of $\Gamma_{QSO} \sim 10^{-12}$ s^{-1} is implied (see § 11.3.3), resulting in a typical equilibration timescale of $t_{eq} \sim 3 \times 10^4$ yr. Full ionization equilibrium is reached after a few equilibration times, or for $t_Q \sim 10^5$ yr. The region in which ionization equilibrium is reached, and hence also the size of the quasar's proximity zone, will grow with time until $t_Q \approx 3t_{eq}$. Thus, even for a highly ionized IGM, the sizes of the proximity zones are dependent on the quasar age for $t_Q \lesssim 10^5$ yr, whereas for ages $t_Q \gtrsim 10^5$ yr the proximity zone

sizes are largely independent on the precise quasar age[4].

In Fig. 11.13 we show the dependence of proximity zone size on quasar age determined from radiative transfer simulations in a highly ionized IGM. Quasars with magnitudes $M_{1450} = -25.0$ (left), $M_{1450} = -27.0$ (middle), and $M_{1450} = -29.0$ (right) are shown, where we have simulated 900 proximity zones at each magnitude. The black horizontal dashed lines show the measured sizes of the proximity zones of CFHQS J2229 + 1457 (left panel), SDSS J1335 + 3533 (middle panel) and SDSS J0100 + 2802 (right panel).

The distribution of proximity zone sizes in the left and middle panel encompasses values of $R_p \approx 1$ pMpc that we have measured only when assuming a short quasar age of $t_Q \lesssim 10^5$ yr. Hence a plausible explanation for the exceptionally small sizes of the proximity zones could be young ages for these quasars. The proximity zone of the very bright quasar SDSS J0100 + 2802 (right panel) is not as significant of an outlier as the other two objects, but nevertheless the small size of its zone would be more probable for $t_Q \sim 10^5$ yr.

As discussed in § 5.3, at lower redshifts $z \lesssim 4$, studies of quasar clustering imply that the duty cycle of quasar activity is $t_{dc} \sim 10^9$ yr (Shen et al., 2007; White et al., 2008)[5]. However, whereas clustering constraints the duty cycle, proximity zones actually probe a different timescale, which is the duration of quasar emission episodes which could be considerably shorter. Indeed, we actually obtain a lower limit on the episodic lifetime, because the quasar could continue to emit for many years after we observe it on Earth. Nevertheless, for a given episodic lifetime t_{episodic} the probability of measuring an age of t_Q is $p = t_Q/t_{\mathrm{episodic}}$, assuming the simplest "light-bulb" lightcurve for the quasars. If we assume an average episodic lifetime of $t_{\mathrm{episodic}} \sim 10^8$ yr, the probability of observing a quasar that has only been shining for $\sim 10^5$ yr is $p \sim 0.1\%$, or $p \sim 1\%$ for $t_{\mathrm{episodic}} \sim 10^7$ yr. We have discovered three objects suggesting ages of $t_Q \sim 10^5$ yr in a sample of 30 quasars, i.e. $p \approx 10\%$. To be consistent with finding a few of these small proximity zones, the quasar episodic lifetimes would need to be $t_{\mathrm{episodic}} \sim 10^6$ yr. This would leave the sizes of the proximity zones of the vast majority of quasars unchanged, which we typically observe much later in their evolution, but could explain the very small zones we find.

---

[4]A mild increase in proximity zone sizes for long quasar ages, i.e. $t_Q \gtrsim 10^7$ yr, can be attributed to heating effects due to He II reionization (Bolton et al., 2010, 2012)

[5]Whether this also applies at redshifts $z \sim 6$ still remains to be determined.

11.5 Conclusions for the Growth of Black Holes

Our analysis uncovered three quasars with exceptionally small proximity zones, two of them with $R_p \lesssim 1$ pMpc. The occurrence of zones this small is extremely rare in our simulations. We estimate the probability of finding these objects in a sample of 30 quasar spectra to be $p \approx 1\%$ for SDSSJ1335 + 3533, and $p \approx 3\%$ for CFHQSJ2229 + 1457 and SDSSJ0100 + 2802. We search for evidence of proximate absorption systems such as DLAs or LLSs, or islands of remaining neutral hydrogen associated with patchy reionization, both of which could result in truncation of the proximity zones closer to the quasars. However the absence of strong metal absorption lines or large GP troughs (in both the Lyα and the Lyβ forest) appears to disfavor these truncation scenarios. Nevertheless we cannot completely rule out the presence of low metallicity or metal-free absorbers due to the low signal-to-noise ratio of the data of CFHQSJ2229 + 1457 and SDSSJ1335 + 3533 and will revisit this question in the next chapter.

The most plausible explanation for the small proximity zones that we observe is that these quasars are relatively young. It takes the IGM roughly $\sim 10^5$ yr to reach ionization equilibrium with the quasar radiation. Proximity zones of quasars with ages longer than this timescale are essentially independent of the exact age of the quasar. However, for shorter quasar ages the surrounding IGM has not yet reached ionization equilibrium resulting in proximity zones comparable to the small sizes we observe. If the duration of quasar emission episodes is $t_{\mathrm{episodic}} \sim 10^8$ yr, the detection of these small zones would be very unlikely, i.e. $p \sim 0.1\%$. This discrepancy can be resolved if one assumes a shorter duration of $t_{\mathrm{episodic}} \sim 10^6$ yr, resulting in a probability of $p \sim 10\%$ for finding these objects with ages of $t_Q \sim 10^5$ yr.

However, an average episodic lifetime of $t_{\mathrm{episodic}} \sim 10^6$ yr causes significant tension with the inferred sizes of SMBHs at these redshifts, since the presence of $\sim 10^9 M_\odot$ SMBHs at $z \sim 6$ requires that quasars accrete continuously at around the Eddington limit for nearly the entire Hubble time of $\sim 10^9$ yr (Volonteri, 2012). Thus, although the proximity zone measurements constrain the episodic lifetime t_{episodic}, whereas the sizes of SMBHs depend on the integral over all emission episodes, i.e. the duty cycle, the presence of large SMBHs at $z \sim 6$ either requires long episodic lifetimes comparable to the duty cycle $\sim 10^9$ yr, or that the quiescent time between episodic emission bursts must be very short, implying that SMBHs grow via many short episodic phases (Schawinski et al., 2015; Mao & Kim, 2016).

In the next chapter we will have a detailed look at one of the young quasars

found, i.e. SDSS J1335 + 3533, and conduct the first spectroscopic study to identify potential differences that might distinguish the young objects from the quasar population at large.

First Spectroscopic Study of a Young Quasar

In chapter 11 we showed how measurements of proximity zone sizes provide constraints on quasar lifetimes for quasars at high redshift, i.e. $z \gtrsim 6$, which are essential in order to put constraints on theoretical models for the formation and growth of the first SMBHs. The proximity zone is sensitive to the quasar lifetime, because the IGM has a finite response time to the quasars' radiation, i.e. the gas reaches its new ionization equilibrium state after a timescale $t_{\mathrm{eq}} \approx \Gamma_{\mathrm{HI}}^{-1} \approx 3 \times 10^4$ yr, where Γ_{HI} denotes the photoionization rate (e.g. Khrykin et al., 2016; Eilers et al., 2017a). This dependency of the proximity zone size R_p on quasar lifetime inferred from radiative transfer simulations (see § 11.2) is shown in the left panel of Fig. 12.1.

We analyzed the proximity zones of an ensemble of 34 high redshift quasars in the previous chapter, and identified three objects that exhibit extremely small proximity zones, indicating very short quasar lifetimes of $t_Q \lesssim 10^5$ yr. One of those young quasars SDSS J1335 + 3533 (hereafter J1335 + 3533) at $z = 5.9012 \pm 0.0019$, exhibits a particularly small proximity zone, i.e. $R_p = 0.78 \pm 0.15$ pMpc, shown as the red dashed line in the left panel of Fig. 12.1. The uncertainty on R_p results from the redshift error, which is very small due to the precise systemic redshift estimate from the CO(6 − 5) emission line of the host galaxy (Wang et al., 2010). Given the proximity zone measurement we can estimate the probability distribution for the lifetime of J1335 + 3533 (right panel), which implies a quasar lifetime of only $t_Q = 10^{3.0 \pm 0.8}$ yr (median and 95% confidence level)[1]. Note however, that the estimate on t_Q is dependent on prior assumptions, for which we chose a flat prior within $2 \leq \log t_Q \leq 8$ yr.

---

[1] This estimate implies that by the time we observe this quasar, it is roughly as old as the pyramids in Egypt are today.
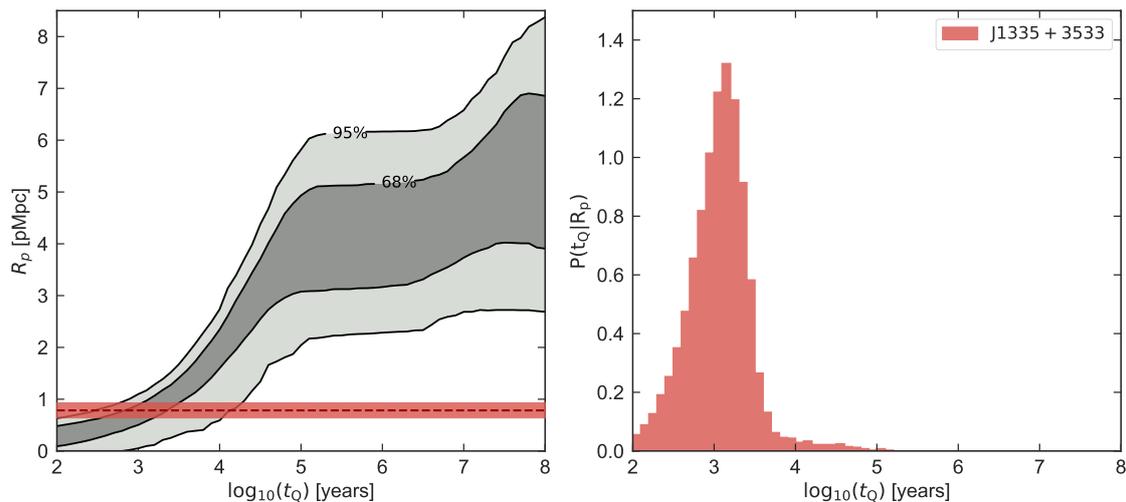
Figure 12.1 *Left panel:* Dependence of the proximity zone size R_p on quasar lifetime t_{Q} from radiative transfer simulations for quasars with similar properties as our young object J1335 + 3533, i.e. $z = 5.9012$ and magnitude $M_{1450} = -26.67$: For short lifetimes, i.e. $t_{\mathrm{Q}} \lesssim t_{\mathrm{eq}}$, the proximity zone grows with lifetime, whereas for longer lifetimes, i.e. $t_{\mathrm{Q}} > t_{\mathrm{eq}}$, the proximity zone "saturates" and ceases to grow. For very long lifetimes of $t_{\mathrm{Q}} \gtrsim 10^{7.5}\,\mathrm{yr}$, the proximity zones grow slightly again due to heating by the reionization of He II. The red dashed line indicates the proximity zone measurement of the young quasar J1335 + 3533, i.e. $R_p = 0.78 \pm 0.15$ pMpc, and its uncertainties (Eilers et al., 2017a). *Right panel:* Probability distribution of t_{Q} for J1335 + 3533 from its proximity zone measurement.

In this chapter we present the first detailed spectroscopic study of a young quasar, i.e. J1335 + 3533. In § 12.1 we present new optical and NIR data on the young quasar. We discuss spectral features of the spectrum, measure its black hole mass, and compare its properties to other co-eval quasars of similar luminosity in § 12.2. We show how our new data rule out any other interpretations of its small proximity zone, such as a premature truncation due to close associated absorption systems in § 12.3. We discuss several scenarios that could potentially explain the rapid growth of the SMBH in the center of J1335 + 3533 in § 12.4, before highlighting a possibility of how we will be able to distinguish between these different scenarios with the upcoming capabilities of JWST .

12.1 New Observations of J1335+3533

In this section we present the new optical (§ 12.1.1) and NIR (§ 12.1.2) data obtained for the young quasar J1335 + 3533.

12.1.1 Optical Spectrum from Keck/DEIMOS

We acquired three exposures of 20 minutes each resulting in a total exposure time of 1 hour on May 26, 2017 (PI: Hennawi), with the Deep Imaging Multi-Object Spectrograph (DEIMOS) at the Nasmyth focus on the Keck II telescope. We used a custom-made slitmask with a $1''$ slit and the 830G grating, resulting in a pixel scale of $\Delta\lambda = 0.47\text{Å}$ ($\Delta v \approx 16\,\text{km}\,\text{s}^{-1}$) and a spectral resolution of $R \approx 2500$. The grating was tilted to a central wavelength of 8100Å such that the wavelength coverage of the spectrum ranges from $6260\text{Å} \leq \lambda_\text{obs} \leq 10080\text{Å}$. We reduced the data using standard techniques with the open source python code PYPIT[2], which is available on GitHub[3].

12.1.2 Near Infra-Red Spectrum from Gemini/GNIRS

We obtained the NIR spectrum of J1335+3533 from the Gemini Observatory Archive[4]. The quasar was observed with the Gemini Near Infra-Red Spectrograph (GNIRS) instrument with the 32 l/mm grating and the short camera's cross-dispersion prism on January 27, 2016, in a program aiming to obtain NIR data for 60 quasars at $z > 5.7$ to study their physical properties, such as black hole masses, metallicities and absorption features (Program: GN-2016A-LP-7, PI: Shen). The data was collected following an ABBA dither pattern using twelve 300 s individual exposures, resulting in a total of 1 hour exposure time. The spectrum covers the wavelength range between $0.85 - 2.5\,\mu\text{m}$ with a resolution of $R \approx 1800$. The data has been reduced making use of the Low-Redux pipeline developed as part of the XIDL suite of astronomical routines in IDL, which employs standard data reduction techniques.

The Mg II emission line of the quasar, which we use to measure its black hole mass (see § 12.2.2), lies in a spectral region with substantial telluric absorption. We followed the telluric standard star correction procedure that is part of the XIDL pipeline in order to minimize the effects of atmospheric absorption. This procedure compares the telluric standard star taken before and after the observation to models of stellar spectra, thus deriving both the telluric correction and the flux calibration simultaneously. We apply the telluric corrections using both standard stars and found that both standard stars yield the same results. As an additional check, we also applied one telluric standard correction to the other standard star and found that our pipeline was producing good corrections in the wavelength range around Mg II .

---

[2] http://pypit.readthedocs.io/en/latest/
[3] https://github.com/PYPIT/PYPIT
[4] https://archive.gemini.edu
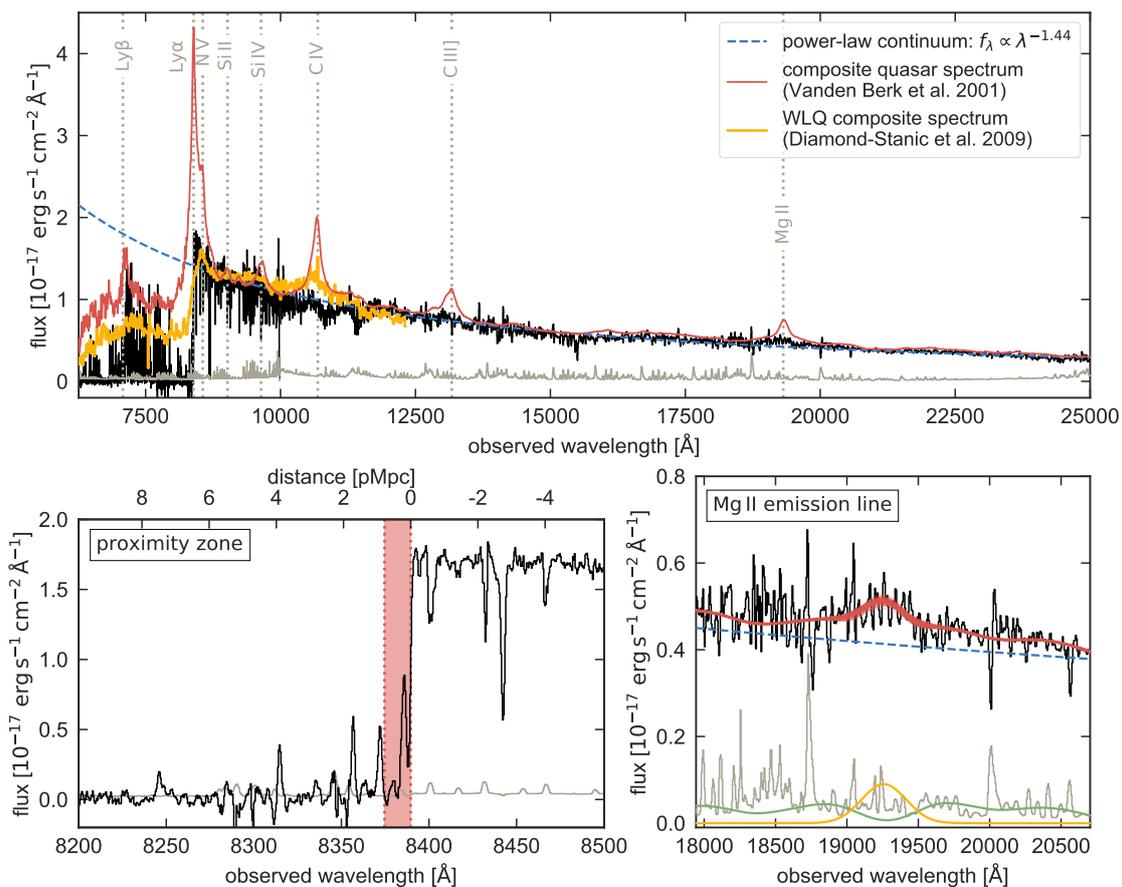
Figure 12.2 *Top panel:* The new optical and NIR spectrum of J1335 + 3533 and its noise vector are shown in black and grey, respectively. A composite quasar spectrum from low-redshift quasars (Vanden Berk et al., 2001) is overplotted in red, a composite spectrum for quasars with weak emission lines is shown in yellow (Diamond-Stanic et al., 2009), as well as a power-law continuum estimate of J1335 + 3533, which is shown as the blue dashed line. *Lower left panel:* Zoom-in into the proximity zone of the quasar indicated by the red shaded area. *Lower right panel:* Zoom-in into the spectral region around the Mg II emission line. The spectrum and its noise vector are shown in black and grey, respectively. The quasar continuum fit and its individual components are shown as the colored curves: a power-law continuum from the nuclear emission of the quasar (blue dashed curve), the smoothed iron template spectrum to mimic the emission from the broad line region (green curve), a Gaussian function to estimate the width of the broad Mg II emission line (yellow curve), and the total continuum estimate (red curve).

Our new optical and NIR data of J1335 + 3533 is shown in the top panel of Fig. 12.2, as well as a zoom-in onto the region around the extremely small proximity zone (lower left panel). We normalized the GNIRS data to match the $J-$band magnitude $J_{\mathrm{AB}} = 19.84$ and re-scaled the DEIMOS spectrum slightly to match the NIR data at $\lambda \approx 10{,}000\text{Å}$.

12.2 Spectral Features

The new data presented in the previous section shows that J1335 + 3533 exhibits very weak emission lines (§ 12.2.1). The NIR spectrum covering the Mg II emission line enables us to estimate the mass of the central SMBH (§ 12.2.2) of J1335 + 3533 and to compare the quasar's properties to other quasars at similar redshift (§ 12.2.3).

12.2.1 Weak Emission Lines

A comparison between the spectrum of J1335 + 3533 and a composite quasar spectrum created from > 2200 quasar spectra at redshifts $0.044 \leq z \leq 4.789$ from SDSS (Vanden Berk et al., 2001) as well as composite spectrum of 32 weak emission line quasars (WLQs) at $z > 3$ (Diamond-Stanic et al., 2009) shown in the top panel of Fig. 12.2, clearly reveals the lack of strong emission lines in the spectrum of J1335 + 3533, as previously noticed by Fan et al. (2006b). We measure an equivalent width (EW) of EW(Lyα + N V) ≈ 5.1Å by fitting two Gaussian functions to the red side of the Lyα + N V emission lines at $\lambda_{\rm rest} \geq 1215.67$Å, which falls into the category of a WLQ according to the definition by Diamond-Stanic et al. (2009), who defined WLQs as quasars with EW(Lyα + N V)< 15.4Å.

The top panel of Fig. 12.2 shows a power-law continuum fit to all wavelengths $\lambda_{\rm rest} \geq 1215.67$Å, with an estimated slope $\alpha_\lambda = -1.444 \pm 0.002$, which is consistent with other quasar spectra (Vanden Berk et al., 2001; Lusso et al., 2015; Diamond-Stanic et al., 2009).

12.2.2 Measurement of the Black Hole Mass

We derive the black hole mass from the single-epoch NIR spectrum covering the Mg II $\lambda 2798.7$Å emission line assuming that the dynamics in the quasar's BLR are dominated by the gravitational pull of the black hole and the virial theorem can be applied. To estimate the black hole mass we use the relation

$$\frac{M_{\rm BH}}{M_\odot} = 10^{6.86} \left(\frac{\rm FWHM_{MgII}}{10^3\,\rm km\,s^{-1}} \right)^2 \left(\frac{\lambda L_{\lambda,3000{\rm Å}}}{10^{44}\,\rm erg\,s^{-1}} \right)^{0.5} \tag{12.1}$$

which has been calibrated by Vestergaard & Osmer (2009) to pre-existing black hole mass scaling relations using the Full Width Half Maximum (FWHM) of other emission lines, such as Hβ or C IV, by means of several thousand high-quality spectra from SDSS.

In order to obtain an estimate of the black hole mass of J1335+3533, we model the quasar emission within the wavelength region around the Mg II line of the quasar, i.e. $2100\,\text{Å} \leq \lambda_{\text{rest}} \leq 3088\,\text{Å}$, as a superposition of a power-law continuum arising from the quasar's nucleus, a scaled template spectrum of the emission from the iron lines, Fe II and Fe III , arising from the BLR of the quasar and a single Gaussian function to model the Mg II emission line, i.e.

$$f_\lambda = f_0 \lambda^{-\alpha_\lambda} + f_1 \cdot f_{\lambda,\text{iron}} + f_2 \cdot \exp\left(-\frac{(\lambda - \mu_{\text{MgII}})^2}{2\sigma_{\text{MgII}}^2}\right)$$

We make use of the iron template spectrum from Vestergaard & Wilkes (2001) and convolve the original template derived from a narrow emission line quasar with a Gaussian kernel to adapt it to the width of the broad emission lines in the spectrum of J1335+3533[5].

We apply the Markov Chain Monte Carlo (MCMC) affine-invariant ensemble sampler emcee[6] (Foreman-Mackey et al., 2013) to fit the quasar spectrum, where our likelihood is $-2\ln\mathscr{L} = \chi^2$ making use of the noise vector of the spectrum. We assume flat priors for the free parameters, i.e. $f_0 \in [0,10]$, $\alpha_\lambda \in [-2,0]$, $f_1 \in [0,1]$, $f_2 \in [0,1]$, $\sigma_{\text{MgII}} \in [0,1000]$, $\mu_{\text{MgII}} \in [-1000,1000]$. We adopt the median of the resulting posterior probability distributions as our parameter estimates. For the slope of the power-law continuum within this small wavelength region around Mg II we obtain $\alpha_\lambda = 1.21 \pm 0.02$, which differs slightly from the slope we inferred for the whole spectrum. From the power-law continuum we can derive the luminosity $\lambda L_{\lambda,3000\,\text{Å}} = (3.15 \pm 0.05) \times 10^{46}\,\text{erg}\,\text{s}^{-1}$.

For the Gaussian fit to the Mg II line we obtain a width of FWHM = $5670 \pm 401\,\text{km}\,\text{s}^{-1}$ and we measure a significant blueshift of $\Delta v_{\text{MgII}} = -870 \pm 179\,\text{km}\,\text{s}^{-1}$ compared to the systemic redshift of the quasar at $z = 5.9012$, derived from measurements of the CO (6-5) 3 mm line by Wang et al. (2010). Such large blueshifts of the Mg II line have been observed in other high redshift quasars (Venemans et al., 2016). The resulting continuum fit around the Mg II emission line and its individual components are shown in the lower right panel of Fig. 12.2. Given Eqn. (12.1) we estimate the mass of the central SMBH of J1335+3533 to be

$$M_{\text{BH}} = (4.13 \pm 0.59) \times 10^9\,M_\odot.$$

Note that the uncertainty in this measurement does not include the systematic intrinsic scatter of 0.55 dex in the scaling relation itself (Vestergaard & Osmer,

---

[5]The width of the Gaussian smoothing kernel $\sigma_{\text{conv}} \approx 2089\,\text{km}\,\text{s}^{-1}$ is determined by Eqn. (1) in Vestergaard & Wilkes (2001), assuming a FWHM $\sim 5000\,\text{km}\,\text{s}^{-1}$ of the quasar's emission lines.

[6]http://dfm.io/emcee/current/
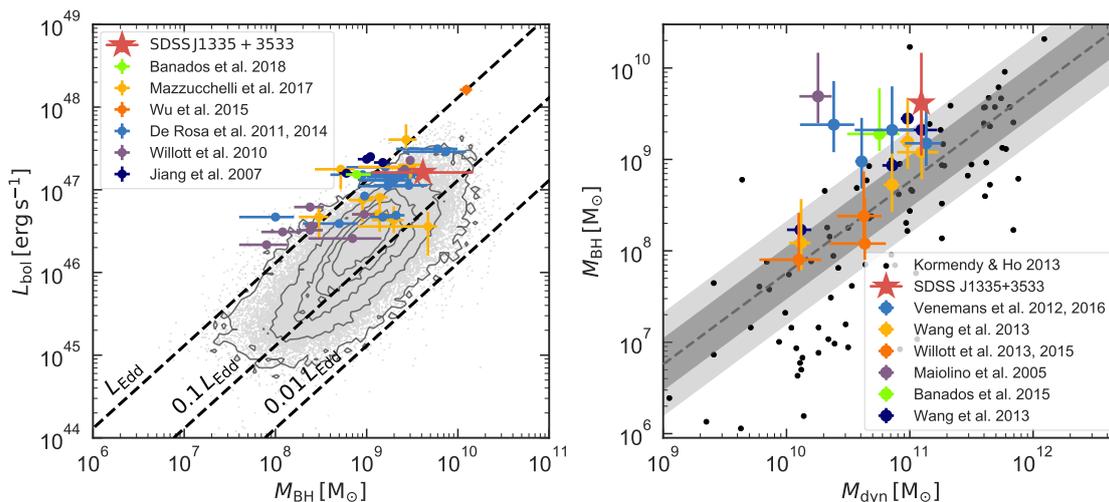
Figure 12.3 *Left panel:* Relation between bolometric luminosity $L_{\rm bol}$ and black hole mass $M_{\rm BH}$ for a sample of low redshift quasars at $0.4 \lesssim z \lesssim 2.2$ from the SDSS Data Release 7 (Shen et al., 2011) indicated by the grey contours and high redshift quasars, which are shown by the colored data points. The young quasar $J1335 + 3533$ is indicated as the red star. Note that the error bars here are smaller than the symbol itself. *Right panel:* Relation between the quasar's black hole mass $M_{\rm BH}$ and its dynamical mass $M_{\rm dyn}$. The grey dashed line shows the relation derived from local galaxies by Kormendy & Ho (2013) with its 1σ and 2σ uncertainties. $J1335 + 3533$ is indicated by the red star, whereas other measurements for high redshift quasars are shown as the colored data points.

2009). Including this systematic uncertainty leads to an estimate of

$$M_{\rm BH} = (4.13^{+10.54}_{-3.02}) \times 10^9 \, M_\odot,$$

which we will continue to use for our analysis in the following section.

12.2.3 Black Hole Mass Scaling Relations

It is interesting to know how the properties of the young quasar $J1335 + 3533$ compare to other co-eval quasars with similar luminosities. We infer the bolometric luminosity for $J1335 + 3533$ following Shen et al. (2008) via $L_{\rm bol} = 5.15 \, \lambda L_{\lambda, 3000\text{Å}} = (1.62 \pm 0.03) \times 10^{47} \, \text{erg s}^{-1}$. From $M_{\rm BH}$ which we estimated in the previous section, we can compute its Eddington luminosity $L_{\rm Edd} = (5.37^{+13.70}_{-3.93}) \times 10^{47} \, \text{erg s}^{-1}$, indicating that the quasar is accreting with an Eddington ratio of $L_{\rm bol}/L_{\rm Edd} = 0.30^{+0.77}_{-0.22}$.

In the left panel of of Fig. 12.3 we show the relation between the bolometric luminosity and the black hole mass of $J1335 + 3533$ in comparison to a low redshift quasar sample of $\gtrsim 75{,}000$ objects at $0.4 \lesssim z \lesssim 2.2$ from the SDSS Data Release 7 (Shen et al., 2011; Wang et al., 2015), and in comparison to several other quasars

at high redshift, i.e. $z \gtrsim 5.8$ (Jiang et al., 2007; Willott et al., 2010; De Rosa et al., 2011, 2014; Wu et al., 2015; Mazzucchelli et al., 2017; Bañados et al., 2018b). The black dashed curves indicate regions with constant Eddington luminosity (Eddington, 1922; Margon & Ostriker, 1973).

We present the scaling relation between black hole mass and dynamical mass of the quasar's host galaxy $M_{\rm dyn}$ in the right panel of Fig. 12.3. The grey dashed line shows the relation derived by Kormendy & Ho (2013) on which local galaxies (shown as black data points) fall with an intrinsic scatter of 0.28 dex.

The dynamical mass of J1335 + 3533 has been derived by Wang et al. (2010) from the FWHM of the CO(6 − 5) line to be $M_{\rm dyn} \sin^2 i \approx 3.1 \times 10^{10}\, M_{\odot}$. The large $5''$ beam of their observations (≈ 29.2 kpc at $z \sim 6$) does not spatially resolve the source and hence they assumed a quasar host size radius of $R \approx 2.5$ kpc. Assuming an inclination angle of $i = 30°$, we estimate $M_{\rm dyn} \approx 1.24 \times 10^{11}\, M_{\odot}$. The colored data points in the right panel of Fig. 12.3 show measurements of other high redshift quasars (Maiolino et al., 2005; Wang et al., 2013; Willott et al., 2013, 2015; Venemans et al., 2012, 2016; Bañados et al., 2015), which show consistently higher black hole masses than the galaxies in the local universe. Note however, that these measurements are highly uncertain since in most cases the quasar's host galaxy is not resolved and its inclination angle is unknown.

In both the Eddington ratio and the $M_{\rm BH} - M_{\rm dyn}$ relation shown in Fig. 12.3 our young object J1335 + 3533, indicated by the red star, does not hold any special position in the parameter space; its black hole mass, bolometric luminosity and dynamical mass are consistent with other quasars at similar redshifts.

12.3 Search for Associated Absorption Systems

In the previous chapter we argued that the small proximity zone of J1335 + 3533 was unlikely to arise due to a truncation from self-shielding associated absorption systems that are optically thick at the Lyman limit, such as DLAs or LLSs which would block the quasar's ionizing radiation. However, this was based on the ESI spectrum of the quasar, which had a much lower S/N than our new data. Hence, we revisit this question using our new data, which enables a thorough search for strong associated absorption systems along the line-of-sight to the quasar.

In order to prematurely truncate the proximity zone a self-shielding absorption system would need to be located within $\lesssim 1000\,{\rm km\,s^{-1}}$ of the quasar, around the edge of the proximity zone, which ends at $z \approx 5.889$. To this end, we search for signatures of strong absorption line systems in the quasar spectrum near the end of
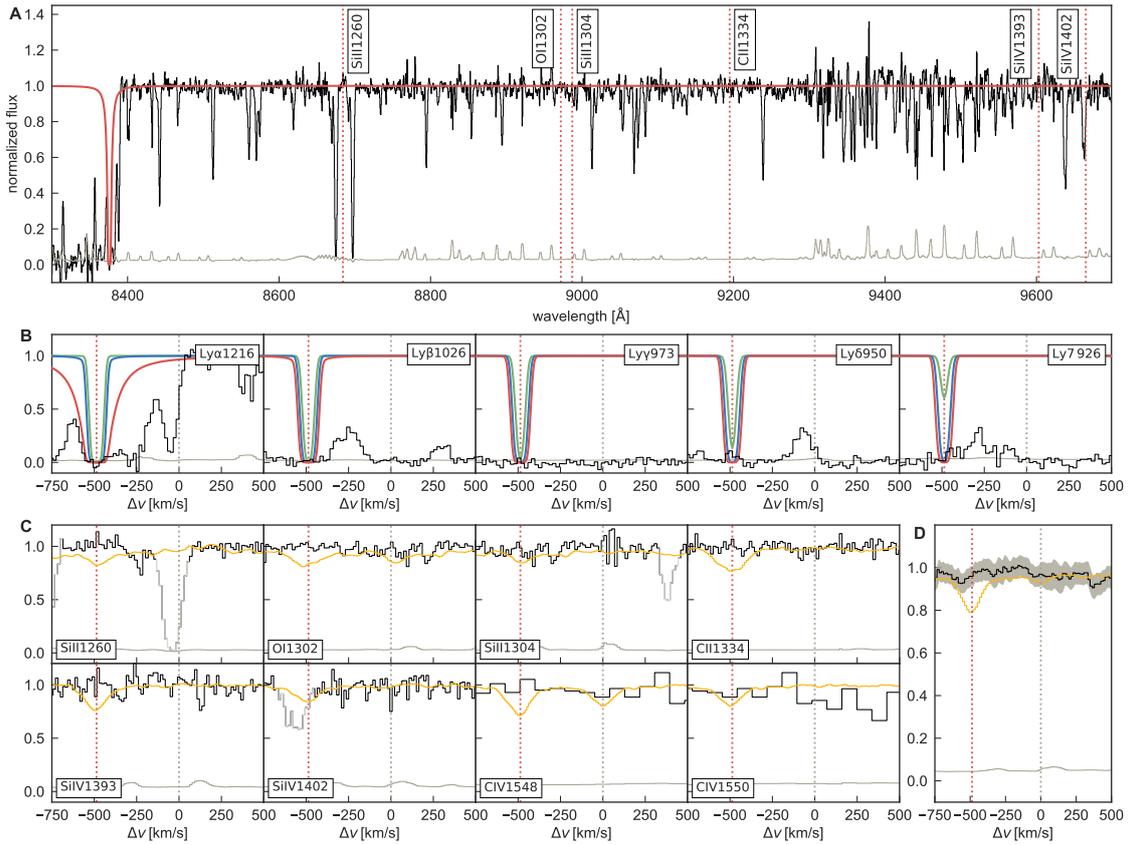
Figure 12.4 *Panel A:* Continuum normalized DEIMOS spectrum of J1335 + 3533 and its noise vector are shown in black and grey, respectively. A possible absorption system ($\log N_{\mathrm{HI}} = 18.5\,\mathrm{cm}^{-2}$, $b = 20\,\mathrm{km\,s}^{-1}$) located in front of the quasar ($z_{\mathrm{abs}} = 5.89$) is shown as the red curve. The expected positions of metal lines that would be associated with the absorption systems are indicated by the red dashed lines. Note that the two strong absorption lines at 8670 Å and 8700 Å are the doublet lines Mg II λ2796 Å and Mg II λ2803 Å that belong to an absorption system at $z_{\mathrm{abs}} \approx 2.102$. *Panels B:* Three absorption systems with varying column densities ($\log N_{\mathrm{HI}} = 18.5\,\mathrm{cm}^{-2}$ (red), $\log N_{\mathrm{HI}} = 17.0\,\mathrm{cm}^{-2}$ (blue), and $\log N_{\mathrm{HI}} = 15.5\,\mathrm{cm}^{-2}$ (green), $b = 20\,\mathrm{km\,s}^{-1}$) smoothed to the spectral resolution of the DEIMOS data within the Lyman series forest. The transmission spikes observed in the Lyδ and Ly7 forest constrain the column density of the potential absorption system. The grey dotted lines mark the position of the quasar. *Panels C:* Zoom-ins into the spectral regions around the expected metal absorption lines. In yellow we show a composite spectrum of 20 LLSs by Fumagalli et al. (2013) centered around the redshift of the absorber (red dashed lines). The gray colored parts of the spectrum mark absorption features from foreground absorption systems. *Panel D:* Stack of the spectral regions around the expected absorption lines. The grey shaded region shows the 1σ uncertainty. The yellow curve shows the stack of the composite LLS spectrum.

its proximity zone by visually inspecting the spectrum, i.e. we search for evidence of damping wings which would indicate the presence of a strong DLA absorber, as well

as associated ionic metal-line transitions, but do not find either. We also consider where an optically thick absorber, i.e. $\log N_{\rm HI} \geq 17\,{\rm cm}^{-2}$, that does not exhibit any damping wings because it might have a small column density and undetected metal absorption lines possibly due to very pristine gas (Fumagalli et al., 2011; Simcoe et al., 2012; Cooper et al., 2015; Cooke et al., 2017), could be located by inspecting its location within the Lyman series forests, since the spectrum at the location of an optically thick absorber should be line black everywhere.

In Fig. 12.4 we show a hypothetical absorption system at $z_{\rm abs} = 5.89$ that would be able to truncate the proximity zone and indicate the locations where metal absorption lines associated with the absorber would fall in the quasar spectrum (panel A). The redshift of the hypothetical absorption system was chosen at the only position possible in the spectrum, such that it lies just at the end of the proximity zone and does not interfere with the clear transmission spikes observed in the Lyα and Lyβ forests. We show the individual regions around the expected metal absorption lines (panels C) and overplot a composite spectrum of 20 LLSs with $\log N_{\rm HI} < 19\,{\rm cm}^{-2}$ by Fumagalli et al. (2013), but do not find any evidence for the presence of such metal absorption lines in the spectrum that could be associated with a close absorption system, neither in the individual spectral regions nor when stacking the spectral regions around the absorption system (panel D).

Additionally, we examine the location of the absorption system within the Lyman series forests (panels B) and find transmission spikes in the Lyδ and Ly7 forests which set a strong limit on the column density of any potential absorption system. In order to not violate the forest transmission, we find that the column density of an absorption system would have to be $N_{\rm HI} \lesssim 15.5\,{\rm cm}^{-2}$ (green curves), which would have an optical depth to ionizing photons of $\tau_{\rm HI} \ll 1$, which is below the threshold of an optically thick absorber, such as a LLS. Hence we conclude that the proximity zone of J1335 + 3533 has not been prematurely truncated by an optically thick absorber.

12.4 Possible Black Hole Formation Scenarios and Prospects for JWST

In this chapter we presented new optical and NIR spectra of the quasar J1335 + 3533, which we previously identified to be a very young object, i.e. $t_{\rm Q} = 10^{3.0 \pm 0.8}$ yr, by analyzing the extent of its proximity zone. With our new high S/N data we can rule out any premature truncation of the proximity zone due to a self-shielding absorption system and constrain the column density of a potential absorber to be

$\log N_{\mathrm{HI}} \lesssim 15.5\,\mathrm{cm}^{-2}$ due to the presence of flux transmission spikes in the higher order Lyman series forest.

By measuring the FWHM of the single-epoch Mg II emission line of the quasar we estimate the mass of the central SMBH to be $M_{\mathrm{BH}} = (4.13^{+10.54}_{-3.02}) \times 10^9 M_\odot$, which is in good agreement with other SMBH mass measurements of co-eval quasars. Other properties of the young quasar and its host galaxy such as the bolometric luminosity, the dynamical mass, the Eddington ratio, and the slope of its continuum emission are likewise in good agreement with other high redshift quasars.

Broadly speaking, J1335 + 3533 is fully consistent with a typical $z \sim 6$ quasar and shows no evidence for anomalies associated with its young age, with the potential exception of the fact that it is a WLQ. Shemmer et al. (2010) discussed whether WLQs could be extreme objects with very high accretion rates, which could potentially explain a younger age. However, the Eddington ratios of WLQs based on the width of their Hβ emission lines that they estimated were not particularly high, i.e. $L_{\mathrm{bol}}/L_{\mathrm{Edd}} \sim 0.4$. Additionally, a study of the X-ray emission of a sample of 11 WLQs with *Chandra* to provide diagnostics of the central accretion process did not reveal any evidence for a hard X-ray spectrum that would be a characteristic of high accretion rates (Shemmer et al., 2009). Interestingly, J1335 + 3533 is a member of the sample, for which upper limits of < 3.0 and < 6.4 counts in the ultrasoft $(0.3 - 0.5~\mathrm{keV})$ and hard X-ray band $(2 - 8~\mathrm{keV})$ were reported, respectively, and $3.0^{+2.9}_{-1.6}$ counts in the soft X-ray band $(0.5 - 2~\mathrm{keV})$ in a 23.47 ks exposure, suggesting only very modest accretion rates.

Hryniewicz et al. (2010) also analyze a WLQ at lower redshift and argue that despite its fairly low Eddington ratio and line properties that are not consistent with the trends expected for quasars with high accretion rates, the most plausible explanation for its weak emission lines is that the quasar activity has just started and the regions from where broad lines are emitted, did not yet respond to this activity.

Following the definition for WLQs by Diamond-Stanic et al. (2009), i.e. EW(Lyα + N V)$< 15.4\,\text{Å}$, Bañados et al. (2016) estimated that 13.7% of the 117 PS1 quasars at $z > 5.7$ that they analyzed fall in this category, which is higher than the fraction of young quasars we estimated. However, out of the three young objects identified in the previous chapter not only J1335 + 3533 but also another object, SDSS J0100 + 2802, exhibits weak emission lines according to (Wu et al., 2015), although Si IV and C IV emission lines are present, suggesting a potentially higher fraction of WLQs among the population of young quasars than in the average quasar population.

The measurement of a billion solar mass black hole in the center of J1335+3533 poses the question of how the SMBH could have grown to be so massive within a time frame of only $t_Q \lesssim 10^4$ yr, when usually at these redshifts quasar lifetimes comparable to the Hubble time with Eddington limited accretion rates have to be invoked in order to explain the existence of SMBHs.

Assuming a simple light-bulb light curve model, which implies that the quasar shines continuously at the same observed luminosity, and a fiducial quasar lifetime of $t_Q \sim 10^8$ yr, the probability of finding a quasar that turned on just $\sim 10^4$ years ago would be very small, i.e. $\sim 0.01\%$. This discrepancy can be resolved if one assumes a shorter fiducial quasar lifetime of $t_Q \sim 10^6$ yr, resulting in a probability of $\sim 1\%$ for finding these objects with short lifetimes, which is roughly consistent with the detection rate in our ensemble. In the previous chapter we identified one quasar, J1335+3533, with $t_Q \lesssim 10^4$ yr, i.e. $1/34 \approx 3\%$, and two additional objects with $t_Q \lesssim 10^5$ yr, i.e. $3/34 \approx 9\%$. However, an average quasar lifetime of $t_Q \sim 10^6$ yr causes significant tension with the inferred masses of SMBHs at these redshifts, since the presence of $\sim 10^9\,M_\odot$ SMBHs at $z \sim 6$ requires that quasars to shine continuously for nearly the entire Hubble time, i.e. $\sim 10^8 - 10^9$ yr (Volonteri, 2010, 2012). Hence, under the assumption of a light-bulb light curve we have to invoke very massive initial seeds ($M_{seed} \sim 10^9\,M_\odot$), for which there is currently no theoretical foundation nor any potential progenitor objects known, or significantly higher, i.e. super-critical, accretion rates (Du et al., 2015; Oogi et al., 2017), which would however be in contrast to the low accretion rates inferred from X-ray observations, in order to explain the growth of the central billion solar mass black hole within the short quasar lifetime.

However, there are several potential scenarios that could explain our findings. One option would be that the quasar has been growing for a much longer time than its lifetime estimate derived from the size of its proximity zone suggests, but it has been growing in a highly obscured phase, such that its UV radiation (and hence the ionizing continuum) has only "broken out" of this obscuring medium $\sim 10^4$ yr ago (e.g. Hopkins et al., 2005; DiPompeo et al., 2017; Mitra et al., 2018). In accordance with this scenario Sanders et al. (1988) proposed that ultraluminous infrared galaxies are the initial, heavily obscured stages of a quasar, which are revealed in the optical as an unobscured quasar at the end of the incipient dust-enshrouded phase. A model proposed by Liu & Zhang (2011) even predicts that the quasars that have just become unobscured should exhibit no broad emission lines.

If this scenario is correct, and obscured quasars are an evolutionary rather than a geometric phenomenon, then the short UV luminous quasar phase which would
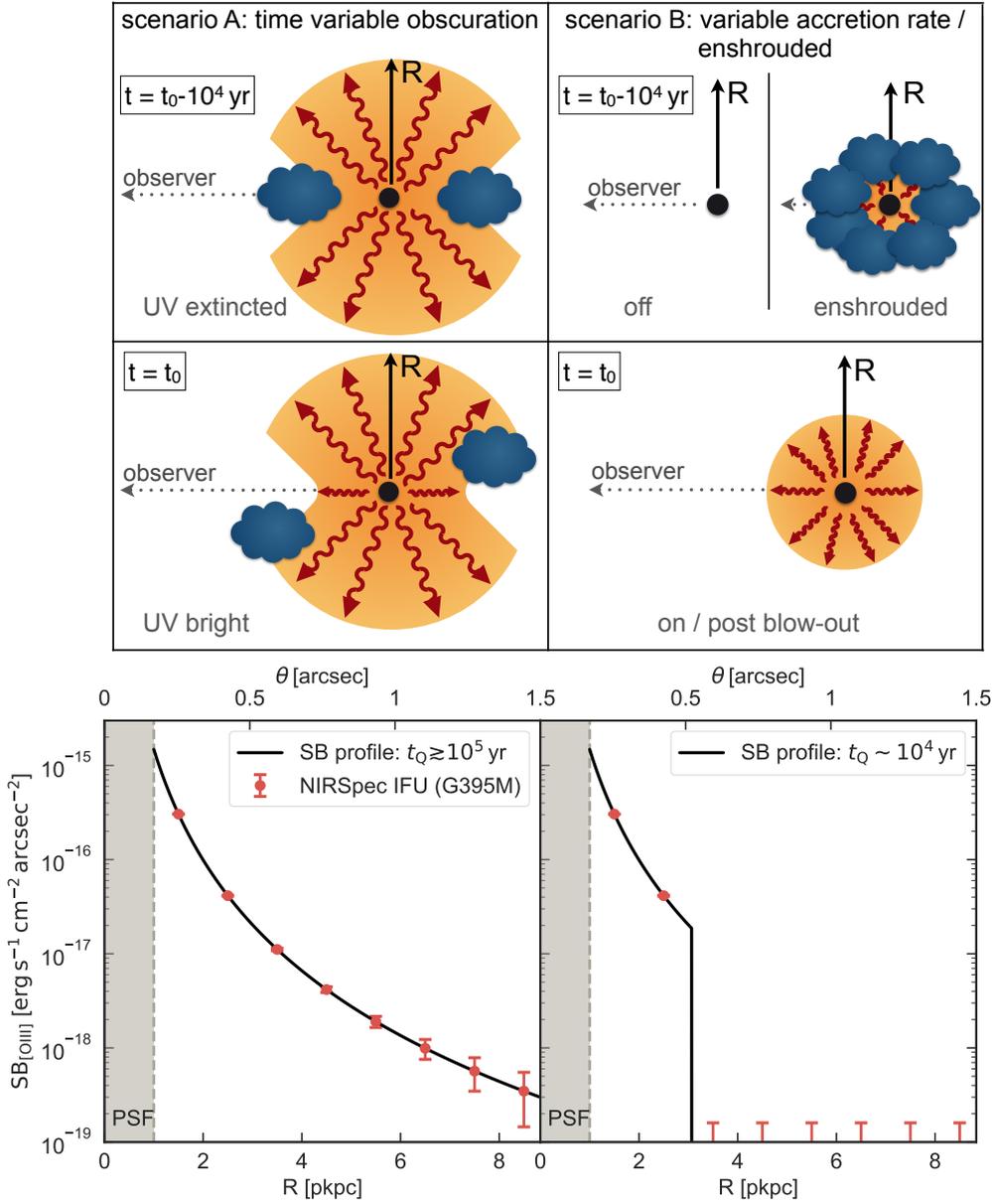
Figure 12.5 *Top panels:* Different black hole growth scenarios that could explain our observations of the young quasar. The left panels show a quasar with a long lifetime, i.e. $t_Q \gtrsim 10^5$ yr, that was previously (at a time $t \leq t_0 - 10^4$ yr) UV obscured along our line-of-sight and is now, at $t = t_0$, UV bright. The right panels show a quasar, whose UV luminous phase just turned on $\sim 10^4$ yr ago, possibly after an initial enshrouded phase. *Bottom panels:* Surface brightness profiles of the ionized extended narrow line region traced by [O III] we expect to observe around the quasar for the two different scenarios. The red data points show the expected signal within a ~ 3 hour exposure with NIRSpec IFU on JWST, when averaged over all spatial pixels in annuli of $\Delta R = 1$ kpc around the quasar and assuming ~ 3 spectral pixels within the [O III] $\lambda 5007$Å emission line when observed with the medium resolution grating G395M.

be consistent with the small proximity zone of J1335 + 3533, implies a very large population of obscured quasars present in the high redshift universe: assuming a fiducial quasar lifetime of at least $t_{\rm Q,fiducial} \sim 10^8$ yr to grow their SMBHs, the inferred UV luminous, i.e. unobscured, quasar lifetime of $t_{\rm Q} \sim 10^6$ yr from our proximity zone measurements, leads to an obscured quasar population that is at least ~ 100 times larger than the known quasar population, which would not have been discovered yet by any survey. This result would be in strong contrast to the estimated fraction of obscured quasars, which is much lower and indicates that there are only $1 - 2$ times as many obscured quasars as unobscured ones (Polletta et al., 2008).

Instead of a light-bulb light curve model we could also assume a flickering light curve model in which the UV continuum emission of the quasar is allowed to vary, which has been suggested for instance in the context of changing-look quasars (e.g. LaMassa et al., 2015). The IGM can probe quasar variability on timescales comparable to the equilibration timescale of the gas, i.e. $t_{\rm eq} \approx \Gamma_{\rm HI}^{-1} \approx 3 \times 10^4$ yr, because of its finite response time to the quasar's radiation (see § 6.3). In a flickering quasar model the variable light curves could either result from intrinsic variations in the accretion rates, in which case active quasar phases would alternate with quiescent galaxy phases, or from variable obscuration by clouds in the BLR passing through our line-of-sight. Assuming an intrinsic variability of quasars due to changes in the accretion rate, our findings would imply episodic quasar lifetimes of $t_{\rm episodic} \sim 10^6$ yr to explain the occasional young quasars with small proximity zones, since the proximity zone measurement is only sensitive to the last active quasar phase, and any evidence of previously active phases in which the SMBH could have grown would be lost. However, a flickering quasar model will also reduce the average proximity zone size (Davies et al. in prep.), which is currently very well reproduced by a light-bulb model.

In order to distinguish between these different black hole formation scenarios we present an idea based on observations of nebular emission lines (such as Hα, Hβ, or [O III] λ5007Å) from ionized gas in the ENLR with an integral field spectrograph. In the aforementioned scenarios, one scenario (scenario B, right panels of Fig. 12.5) predicts that the quasar's UV luminous phase just turned on $\sim 10^4$ yr ago, possibly after an initial, totally obscuring, dust-enshrouded phase, or – under the assumption of a flickering quasar light curve – it could have also turned on again after a quiescent phase. In the other scenario (scenario A, right panels of Fig. 12.5) the quasar has been UV obscured only along our line-of-sight due to clouds and variations in the BLR, but has indeed been shining in other directions

and growing its SMBH for a much longer time.

For either case we would expect to see a different surface brightness profile: If the quasar has indeed been UV obscured along our sight line, but emitting UV ionizing radiation into other directions for a very long time, i.e. $t_Q \gtrsim 10^5\,\mathrm{yr}$, we would expect to see an extended surface brightness profile. However, if the quasar's UV luminous phase turned on (again) only $t_Q \sim 10^4\,\mathrm{yr}$ ago and a flickering quasar model, a dust-enshrouded initial stage, or possible super-critical accretion rates could explain the rapid growth of its SMBH, we expect to observe a surface brightness profile of its ENLR that is truncated at a distance of $R \sim c \times 10^4\,\mathrm{yr} \approx 3$ proper kpc (pkpc), which corresponds to the distance the quasar's radiation could have traveled during its lifetime.

For calculating the two different surface brightness profiles we assumed the typical surface brightness of [O III] around a quasar in the local universe from (Liu et al., 2014) and re-scaled it by $(1+z)^{-4}$ to account for cosmological dimming. We choose the [O III] $\lambda5007\text{Å}$ emission line doublet as a suitable tracer of the ionized ENLR, because it can be observed with NIRSpec on JWST at $z \sim 6$ and traces relatively dense regions $n_H \sim 10^3 - 10^5\,\mathrm{cm}^{-3}$. Thus it will be almost instantaneously collisionally excited once the gas is photoionized to O^{2+}. If the quasar shuts off the timescale for O III to recombine is relatively short, i.e. $t_{rec} \sim 10 - 100\,\mathrm{yr}$[7] and hence even in the case of intermittent quasar activity the [O III] emission from a previously active phase will have faded away if the quasar has been inactive for $t_{off} > t_{rec}$. Thus the two aforementioned scenarios A and B will produce distinct surface brightness profiles.

This chapter presents a new step towards understanding the formation and growth of the first SMBHs in the universe. Our recent discovery of a young quasar population poses an interesting challenge for the current picture of SMBH growth and has put more stringent constraints on current quasar and galaxy evolution theories. Our proposed observations of the ENLR with the upcoming capabilities of JWST promise to shed further light onto the growth rate and formation of the SMBH in the center of the quasars and potentially provide a window to new physics (if super-critical accretion rates need to be invoked to explain the rapid growth), new insights into quasar-galaxy co-evolution (if indeed a flickering in the accretion rate of the quasar could explain our findings), or might indicate a large new population of obscured quasars in the high-redshift universe.

---

[7]The recombination time scale is $t_{rec} \sim \frac{1}{\alpha n_e}$, where $\alpha = 1.99 \times 10^{-12}\,\mathrm{cm}^3\,\mathrm{s}^{-1}$ is the recombination rate coefficient for O III to recombine to 0 II (Draine, 2011), and $n_e \sim 10^3 - 10^5\,\mathrm{cm}^{-3}$ is the electron density.

Part III

Structure and Dynamics of the Milky Way

A Data-Driven Model to Infer Spectrophotometric Parallaxes

This chapter summarizes the results of Hogg, Eilers, & Rix (2018), who developed a data-driven model to estimate precise parallaxes to luminous red giant stars, using only their spectroscopic and photometric information. We will use the results of this study in Chapter 14 to estimate the circular velocity curve of the Milky Way over large Galactocentric distances.

One of the main motivations for large spectroscopic stellar surveys like *APOGEE* (Majewski et al., 2017) or the upcoming *SDSS-V* survey (Kollmeier et al., 2017) is to make precise maps of our home galaxy, the Milky Way, out to large heliocentric radii, even beyond the Galactic center (see § 3.4 for details). These surveys target in particular luminous red giant stars, because those stars are very common and can be observed out to large distances due to their brightness.

In addition to the aforementioned ground-based spectroscopic surveys, the *Gaia* mission (Gaia Collaboration et al., 2016, 2018) funded by the European Space Agency has revolutionized our knowledge about the Milky Way and has ushered us in a new era of Galactic astronomy. However, while *Gaia* provides good measurements of positions and proper motions, it does not deliver precise measurements of parallaxes and thus beyond a few kpc distance from the Sun, uncertainties in the distance estimates dominate everything.

In Hogg, Eilers, & Rix (2018) we developed a purely data-driven model to predict distances to luminous red giant stars by means of their spectroscopic and photometric features, which we will summarize here. In § 13.1 we state the assumptions we make to develop the model, which is introduced in § 13.2. The data set of red giant branch (RGB) stars is presented in § 13.3. We illustrate the results of our new model in § 13.4, before summarizing the results in § 13.5.

13.1 Assumptions of the Method

Our method assumes that red giant stars are dust-correctable, standardizable candles, which means that we can infer their distance modulus – and thus their parallax – from their spectroscopic and photometric features alone. Hence, we use *Gaia* parallax measurements to learn relationships between the spectroscopic and photometric properties of stars and their parallaxes, and afterwards apply that learned model to estimate improved parallaxes for stars with poor (or no) *Gaia* measurements.

Our model further assumes that the logarithmic parallax can be predicted by a completely *linear* function of the chosen features. This linearity is only assumed in a small range of the stellar parameter space, since we restrict our analysis to stars at the top of the red giant branch only, but this assumption limits the flexibility or capacity of the model.

For further details on the assumptions of our model we refer the reader to Hogg, Eilers, & Rix (2018).

13.2 The Data-Driven Model

For each star n we create a D-dimensional feature vector x_n that contains the chosen stellar features, i.e. the logarithm of the continuum-normalized flux (~ 7400 pixels) and the photometric data in multiple bands. The model assumes that the true parallax can be expressed as a linear combination of the feature vector, i.e.

$$\varpi_n^{(a)} = \exp(\theta \cdot x_n) + \text{noise}, \tag{13.1}$$

where $\varpi_n^{(a)}$ is the astrometric parallax measurement from *Gaia*, and θ is the D-dimensional vector that we optimize for, containing the linear coefficients of our model. The log-likelihood function can be expressed as

$$\ln \mathscr{L}(\theta) = -\frac{1}{2}\chi^2(\theta) \quad \text{with} \quad \chi^2(\theta) \equiv \sum_{n=1}^{N} \frac{[\varpi_n^{(a)} - \exp(\theta \cdot x_n)]^2}{\sigma_n^{(a)2}}, \tag{13.2}$$

where $\sigma_n^{(a)}$ is the uncertainty on the astrometric parallax measured by *Gaia*.

Additionally, we expect that only a small subset of the *APOGEE* spectral pixels will be relevant to the prediction of parallax, and thus we add a L1-norm $||x||_1^1$ to the objective function, i.e.

$$\hat{\theta} \quad \leftarrow \quad \underset{\theta}{\text{argmin}} \left[\frac{1}{2}\chi^2(\theta) + \lambda\,||P \cdot \theta||_1^1 \right], \tag{13.3}$$

where λ is the regularization parameter that we set via cross-validation, and P is a projection operator that selects the relevant components. Note that we only add the L1-norm to the spectral pixels of the feature vector, but not to the photometric bands. In the end this norm sets approximately 75% of the pixels to zero.

In order to optimize the model and avoid local optima, we run a cascade of optimizations by training first on the highest signal-to-noise stars and use these results as an initial guess for the full optimization. Once the model is optimized for a training set, the output allows us to predict the so-called *spectrophotometric parallax* $\varpi^{(\text{sp})}$ for all stars in a disjoint validation set. We assign the spectrophotometric parallax according to

$$\varpi_m^{(\text{sp})} \quad \leftarrow \quad \exp(\hat{\theta} \cdot x_m), \tag{13.4}$$

where $\hat{\theta}$ is the optimized coefficient vector and x_m is the feature vector for star m.

Due to the linearity of the model we can easily propagate the uncertainty on the spectrophotometric parallax $\sigma_m^{(\text{sp})}$, i.e.

$$\sigma_m^{(\text{sp})2} \quad \leftarrow \quad \varpi_m^{(\text{sp})2} \hat{\theta}^T \cdot C_m \cdot \hat{\theta}, \tag{13.5}$$

where C_m is a diagonal matrix with the variances of the features of x_m along the diagonal.

13.3 Data Set of Red Giant Stars

Since our model is purely linear, we restrict ourselves to a small region in the stellar parameter space and focus our analysis on stars on the upper end of the red giant branch, i.e. $0 \leq \log g \leq 2.2$, which is more luminous than the red clump. For this selection we make use of the surface gravity $\log g$ estimates inferred by the *APOGEE* pipeline (García Pérez et al., 2016).

For each star we take the *APOGEE* spectral data, i.e. 7405 pixels, on a common rest-frame wavelength grid, which we pseudo-continuum normalize according to the procedure described in Ness et al. (2015). We take the photometric data in the J, H, and K band from *2MASS* (Skrutskie et al., 2006). We use the Gaia Data Archive[1] to match (Marrese et al., 2018) our data set to the *WISE* Catalog (Wright et al., 2010) and obtain mid-IR photometric data in the W_1 and W_2 bands, i.e. at 3.6μm and 4.5μm, respectively. Additionally, we match to the *Gaia* DR2 Catalog (Gaia Collaboration et al., 2018) and add the photometric information in the G, G_{BP}, and G_{RP} bands.

---

[1]https://gea.esac.esa.int/archive/
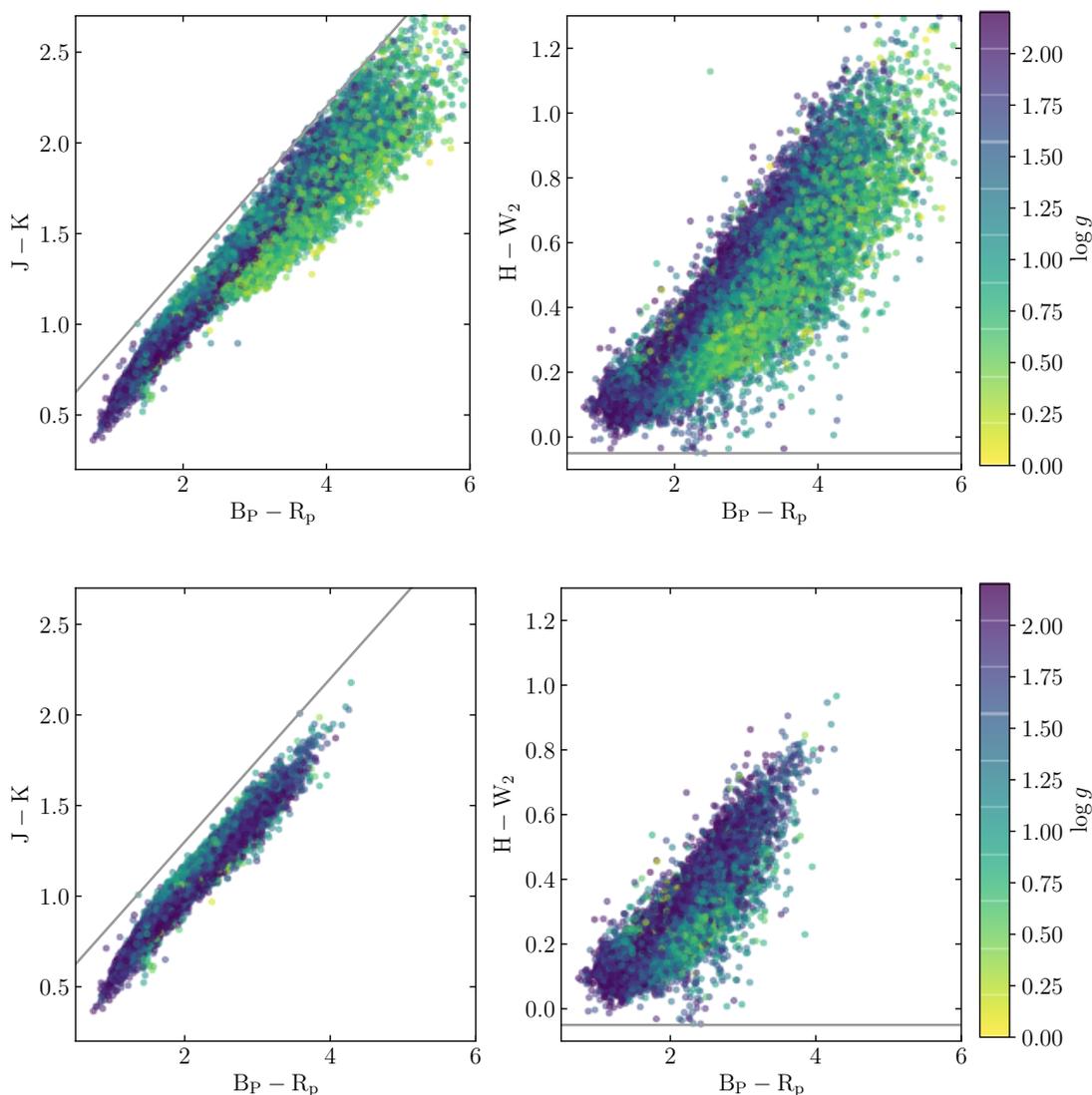
Figure 13.1 Color-color diagrams for all stars in the parent sample (*top*), as well as the training sample (*bottom*), color coded by the surface gravity determined by *APOGEE*. Grey lines indicate two light color cuts to remove stars with contaminated or outlying photometry (see Hogg, Eilers, & Rix, 2018, for details).

Note that we require a complete set of features, i.e. full spectral and photometric data, for each star, as well as a measurement of its astrometric parallax by *Gaia*. This data set, which we call the "parent sample", contains 44,784 stars and is shown in the top panels of Fig. 13.1. For each star we can construct its 7414-dimensional feature vector

$$x_n^T \equiv \left[1, G_n, G_{\text{BP},n}, G_{\text{RP},n}, J_n, H_n, K_n, W_{1,n}, W_{2,n}, \ln f_{1,n}, \ln f_{2,n}, ..., \ln f_{L,n}\right], \quad (13.6)$$

where 1 allows for a linear offset, and $\ln f_{i,n}$ denotes the logarithm of the flux in each pixel i.
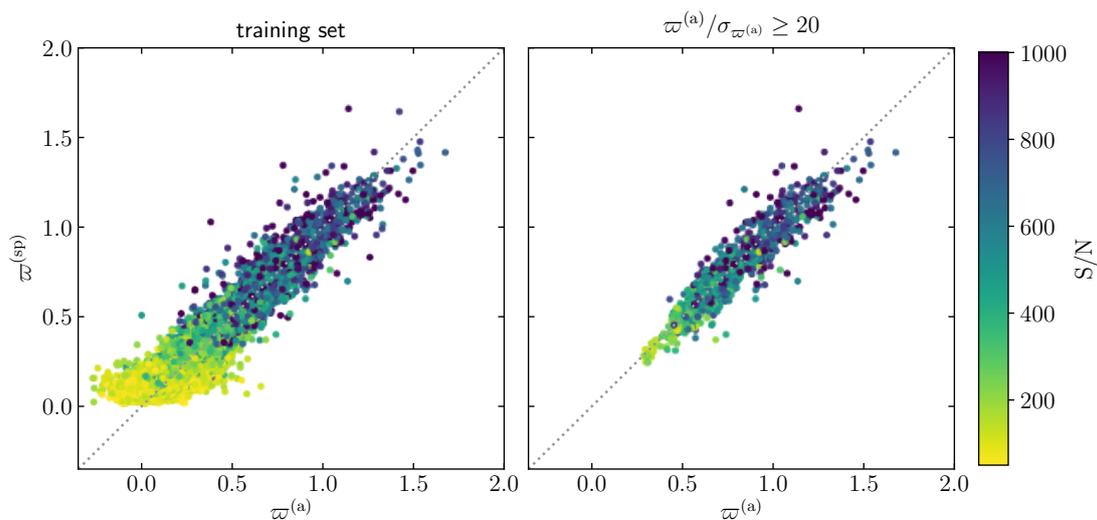
Figure 13.2 Accuracy of our data-driven model in the two-fold cross-validation for our training set (*left*) and stars with the highest signal-to-noise astrometric parallax measurements (*right*), which we use to assess the prediction precision, i.e. $\sim 9-10\%$. The x-axis shows *Gaia*'s measurements $\varpi^{(\mathrm{a})}$, whereas the y-axis denotes the predicted $\varpi^{(\mathrm{sp})}$ measurements. By construction our method does not predict any negative parallax values.

We then define a subset of the parent sample as our training set that fulfills additional quality criteria, such as $\sigma_n^{(\mathrm{a})} < 0.1$ mas, as well as further criteria proposed by Bailer-Jones et al. (2018). The training sample, which consists out of $28,226$ stars, is shown in the bottom panels of Fig. 13.1.

Before optimizing our model we add a positive offset $\varpi_0 = 0.048$ mas to the astrometric parallax measurements to adjust for the slightly under-estimated parallaxes reported by the Gaia Collaboration (Lindegren et al., 2018). We chose the exact value of the offset such that it optimizes our objective function in cross-validation.

13.4 Results and Validation

For the optimization of our model we split our data set randomly into two distinct sets A and B, and optimize our model on training set A and infer the spectrophotometric parallaxes for all stars in the parent sample in part B and vice versa. The results of this cross-validation are shown in Fig. 13.2. For stars with the highest signal-to-noise parallax measurements from *Gaia*, our model predicts spectrophotometric with a scatter of $\sim 9-10\%$.

However, this estimate is based on stars which are likely to be bright and in the
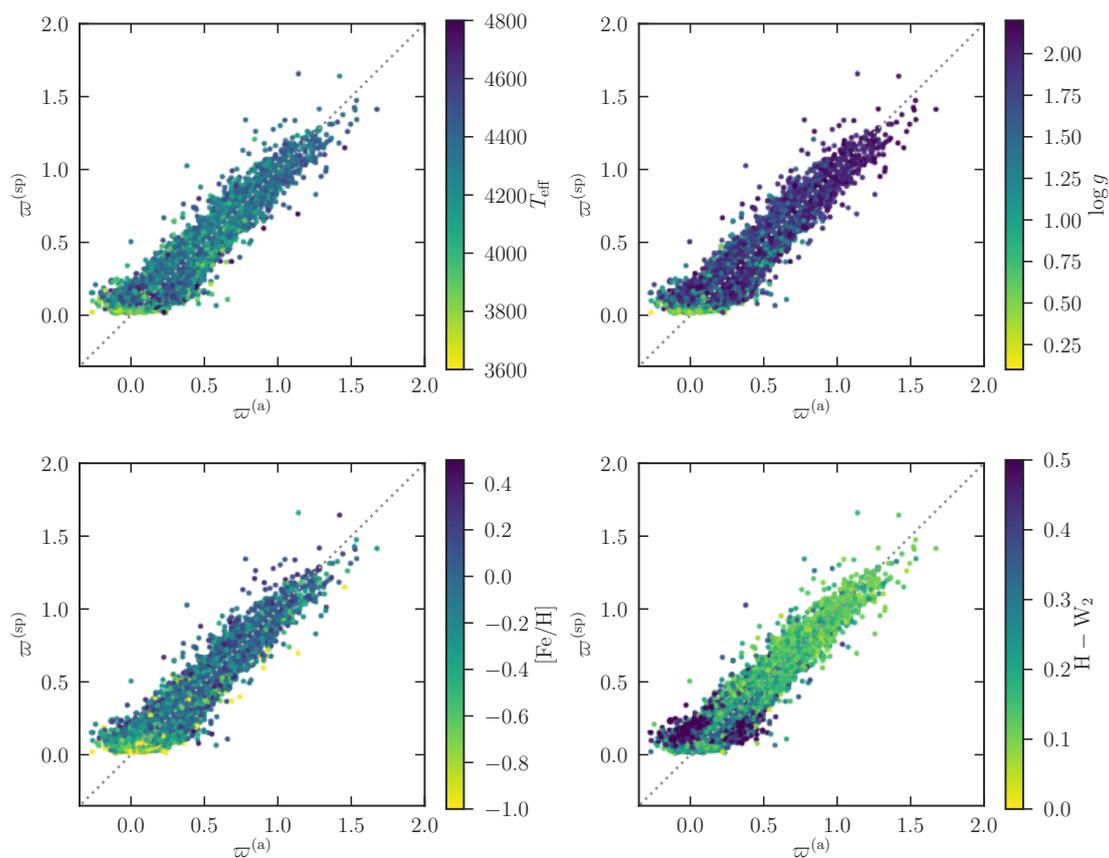
Figure 13.3 Same as Fig. 13.2 but colored by various relevant data features that might influence the residuals of our prediction. However, we do not find any such trends.

Sun's vicinity and thus have very precise astrometric parallax estimates. Hence, in Fig. 13.3 we show whether there are any trends in the precision of our prediction with dust, taking the colors $H - W_2$ as a proxy, or stellar metallicity, but we do not find any indications thereof.

Another validation of our prediction can be obtained by looking at known clusters in our Milky Way (see Fig. 13.4). We compare the two parallax measurements for astrometrically confirmed members, which we identified by means of their celestial coordinates and proper motions, of three clusters that span a range in ages and metallicities. We find that both *Gaia*'s estimates, as well as the predictions from our data-driven model, appear to be unbiased for these clusters, however, with the exception of M71. We are uncertain why M71 gives biased results, and can only guess it might be the result of very high extinction. We have looked at colors, sky position, and cluster properties and have no simple explanation for this discrepancy.
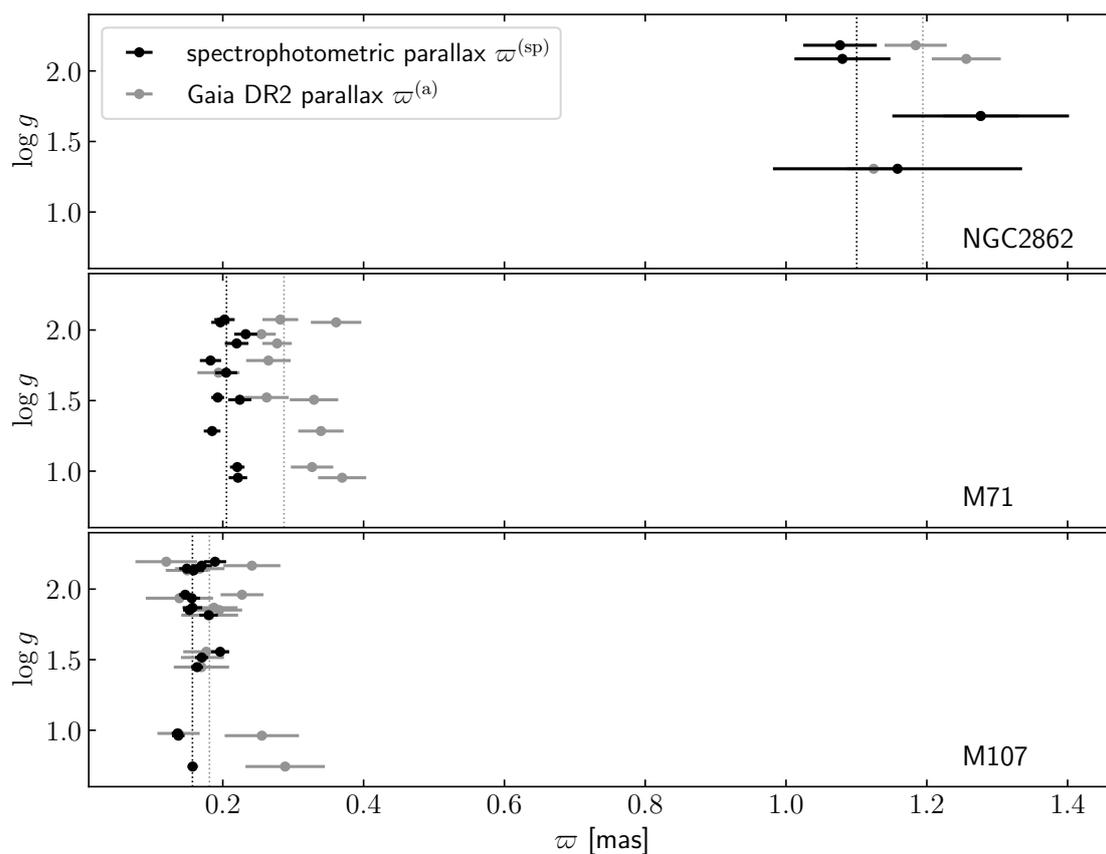
Figure 13.4 Comparison of $\varpi^{(a)}$ and $\varpi^{(sp)}$ for member stars of three stellar clusters, spanning a range of different ages and metallicities.

In Fig. 13.5 we compare the uncertainties we obtain in the spectrophotometric parallaxes and the uncertainties from the astrometric input parallaxes. Our resulting parallaxes are more than twice as precise as *Gaia*'s predictions at heliocentric distances of $\gtrsim 3$ kpc ($\gtrsim 1$ kpc) for stars with $G \sim 12\,\mathrm{mag}$ ($G \sim 14\,\mathrm{mag}$). At ~ 15 kpc distance from the Sun, the derived spectrophotometric distances are a factor of $\approx 6-8$ times more informative and thus they enable us to make precise maps of the Milky Way (see Chapter 14 and Chapter 16 for applications).

The final catalog with our spectrophotometric parallax estimates for all red giant stars in the parent sample has been publicly released via the zenodo platform[2].

13.5 Summary

In this chapter we have shown that we can estimate parallaxes of luminous red giant stars based on their spectroscopy and photometry alone by using a purely lin-

---

[2]https://zenodo.org/record/1468053
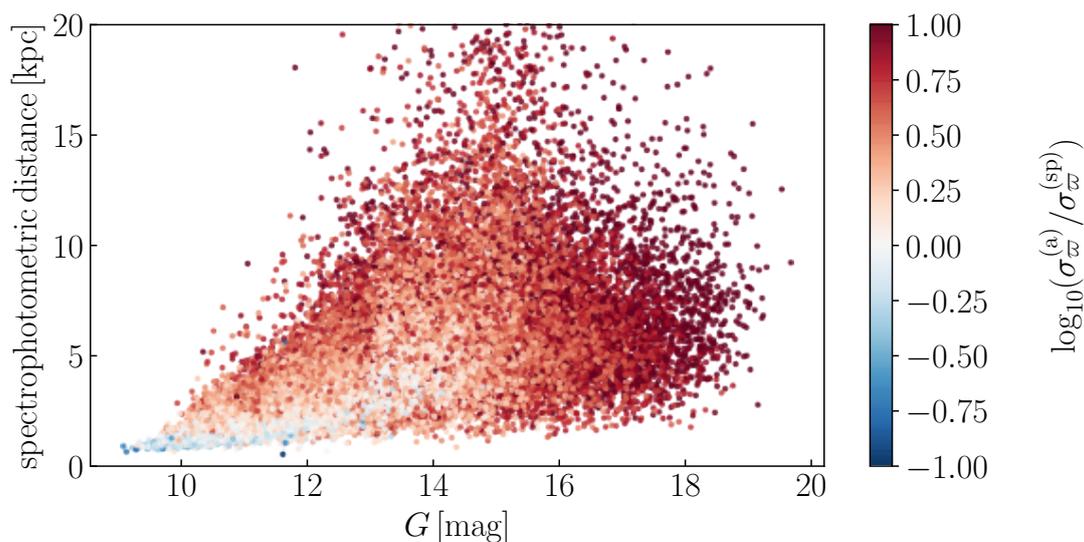
Figure 13.5 Comparison of the parallax uncertainties from the *Gaia*'s astrometric measurements and the predicted spectrophotometric estimate, as a function of the stellar magnitude and its heliocentric distance.

ear data-driven model. Our work assumed that the spectroscopic and photometric properties are sufficient to estimate a dust-independent apparent magnitude and the stellar parameters like age and surface gravity, which are strongly correlated with luminosity.

The design of our model is extremely rigid. We expect a more flexible model that involves for instance Gaussian processes or deep learning techniques to perform better. These complex models can chose the features that are relevant themselves, and go beyond a linear combination of the chosen features. However, these models include the choice and fine-tuning of many hyperparameters, and are computationally expensive. At least in the small range of stellar parameter space that we chose, a linear model seems to be sufficient and has the significant advantage of being computationally extremely fast, as well as easy to interpret.

This work delivered greatly improved distance estimates for luminous red giant stars which we made publicly available. The purpose of this project was to make global maps of the Milky Way. An example of which is shown in Fig. 16.1 in Chapter 16, where we can see the rotation of the disk around the kinematic center of our Galaxy, which we will investigate in more detail in the future. In the following chapter we will use the results of this work to derive the circular velocity curve of the Milky Way.

The Circular Velocity Curve of the Milky Way

In this chapter we present a new measurement of the Milky Way's circular velocity derived from the 6-dimensional phase-space measurements for $\gtrsim 23,000$ red-giant stars at Galactocentric distances between $5 \lesssim R \lesssim 25$ kpc. Critical for such a measurement are precise and accurate distances to luminous tracers that can be observed over a wide range of Galactic distances. Stars at the top of the RGB are frequent and very bright and hence well suited as a tracer population for this study. In chapter 13 we developed a data-driven model making use of spectral data from *APOGEE* DR14 as well as photometric information from *Gaia*, *2MASS*, and *WISE* to improve on the estimated parallaxes from *Gaia* DR2, and derive new parallaxes estimates with only ~ 10% uncertainties.

We derive the Milky Way's circular velocity within the Galactic plane based on the Jeans equation (see § 3.3), which relates the circular velocity to the number density and Galactocentric radial as well as tangential velocity dispersions of the tracer population, assuming an axisymmetric gravitational potential for the Milky Way (see also e.g. Bhattacharjee et al., 2014).

The outline of this chapter is as follows: in § 14.1 we briefly present the data set of red giant stars for which which we previously derived spectrophotometric distances. We introduce the Jeans equation in cylindrical coordinates in § 14.2 and present the resulting rotation curve in § 14.3, before summarizing our results in § 14.4.

14.1 Red Giant Stars within the Stellar Disk

In order to determine the circular velocity curve $v_c(R)$, we need to infer the Galactocentric radius R, the tangential velocity $v_\varphi(R)$, the radial velocity $v_R(R)$, and their uncertainties for each star from the 6-dimensional phase-space information, which requires the knowledge of precise distances to these stars. In chapter 13 we infer parallaxes with a purely linear model, and hence we restrict ourselves to a very limited region of the stellar parameter space. We choose stars on the upper red giant branch with low surface-gravity, i.e. $0 \leq \log g \leq 2.2$, which selects stars that are more luminous than the red clump. Furthermore, we select stars which have existing spectral data from *APOGEE* DR14 (Majewski et al., 2017), as well as complete photometric information in the G-band, G_{BP} and G_{RP} from *Gaia* DR2 (Gaia Collaboration et al., 2016), near-infrared data in the J-, H-, and K-band from *2MASS* (Skrutskie et al., 2006), and $W1$ and $W2$ (at $3.6\,\mu m$ and $4.5\,\mu m$, respectively) from *WISE* (Wright et al., 2010). After applying further data quality cuts (see Hogg, Eilers, & Rix, 2018, for details), this results in $44,784$ luminous red giant stars with complete spectral and photometric data, allowing us to derive precise parallax and hence 6-dimensional phase-space estimates.

For the analysis of the circular velocity curve we only consider stars within a wedge of $60°$ from the Galactic center towards the direction of the Sun, i.e. $\pm 30°$. Additionally, we apply a cut on the height above the Galactic plane in order to restrict ourselves to stars within the disk, i.e. we take all stars which satisfy either $|z| \leq 0.5$ kpc or lie within $6°$ from the Galactic plane, i.e. $|z|/R \leq \tan(6°)$. In order to avoid contamination from halo stars we also exclude stars with a vertical velocity component of $|v_z| > 100\,\mathrm{km\,s^{-1}}$. Furthermore, we select stars with low α-element abundances, i.e. $[\alpha/\mathrm{Fe}] < 0.12$, in order to avoid large asymmetric drift corrections. These cuts results in a total of $23,129$ giant stars for the analysis of $v_c(R)$.

The data-driven, linear model combining only spectroscopic data from *APOGEE* DR14 and photometry from *Gaia* DR2, *2MASS*, and *WISE* which we presented in Chapter 13 provides precise parallax estimates for the selected red giant stars. For our analysis of the circular velocity curve we transform all stars into the Galactocentric coordinate frame making use of the barycentric radial velocities from *APOGEE*, as well as the distance from the Sun to the Galactic center $R_\odot = 8.122 \pm 0.031$ kpc (Gravity Collaboration et al., 2018), its height above the Galactic plane $z_\odot \approx 0.025$ kpc (Jurić et al., 2008), and the Galactocentric velocity of the Sun, i.e. $v_{\odot,x} \approx -11.1\,\mathrm{km\,s^{-1}}$, $v_{\odot,y} \approx 245.8\,\mathrm{km\,s^{-1}}$, and $v_{\odot,z} \approx 7.8\,\mathrm{km\,s^{-1}}$, derived from the proper motion of Sgr A*, which we assume to be the rest-frame of the Galac-

tic center (Reid & Brunthaler, 2004). Note that our method does not require any knowledge of the local standard of rest.

14.2 Jeans Modelling of the Circular Velocity

Assuming an axisymmetric gravitational potential of the Milky Way, we solve the cylindrical form of the Jeans equation (Jeans, 1915), which can be derived from Eqn. (3.2) when applying a Hamiltonian in cylindrical coordinates (see § 4.1 and § 4.8 in Binney & Tremaine, 2008). The Jeans equation links the moments of the velocity distribution and the density of a collective of stars to the gravitational potential, i.e.

$$\frac{\partial \nu \langle v_R^2 \rangle}{\partial R} + \frac{\partial \nu \langle v_R v_z \rangle}{\partial z} + \nu \left(\frac{\langle v_R^2 \rangle - \langle v_\varphi^2 \rangle}{R} + \frac{\partial \Phi}{\partial R} \right) = 0, \qquad (14.1)$$

where ν denotes the density distribution of the tracer population. We can then derive the circular velocity via

$$v_c^2(R) = \langle v_\varphi^2 \rangle - \langle v_R^2 \rangle \left(1 + \frac{\partial \ln \nu}{\partial \ln R} + \frac{\partial \ln \langle v_R^2 \rangle}{\partial \ln R} \right). \qquad (14.2)$$

Note that for deriving Eqn. (14.2) we neglected the second term in Eqn. (14.1), because the cross-term $\langle v_R v_z \rangle$ and its vertical gradient is $\sim 2-3$ orders of magnitude smaller compared to the remaining terms out to Galactocentric distances of $R \sim 18$ kpc and hence effects the circular velocity at the $\sim 1\%$ level (see § 14.3.2).

The terms in Eqn. (14.2) are estimated as follows: In the presence of measurement uncertainties C_v we calculate the velocity tensor V by assigning

$$V \leftarrow \langle v v^T \rangle - C_v, \qquad (14.3)$$

where $\langle v v^T \rangle$ denotes the squared velocity averaged over an ensemble of stars, and replace the respective terms in Eqn. 14.2 with:

$$\langle v_R^2 \rangle \leftarrow V_{RR},$$
$$\langle v_\varphi^2 \rangle \leftarrow V_{\varphi\varphi}.$$

To use Eqn. (14.2) we need to know the radial density profile for the tracer population. Because we do not know the selection function of *APOGEE* precisely enough to determine the profile from the data itself, we simply and plausibly chose an exponential function, i.e.

$$\nu(R) \propto \exp\left(-\frac{R}{R_{\exp}}\right) \qquad (14.4)$$

147
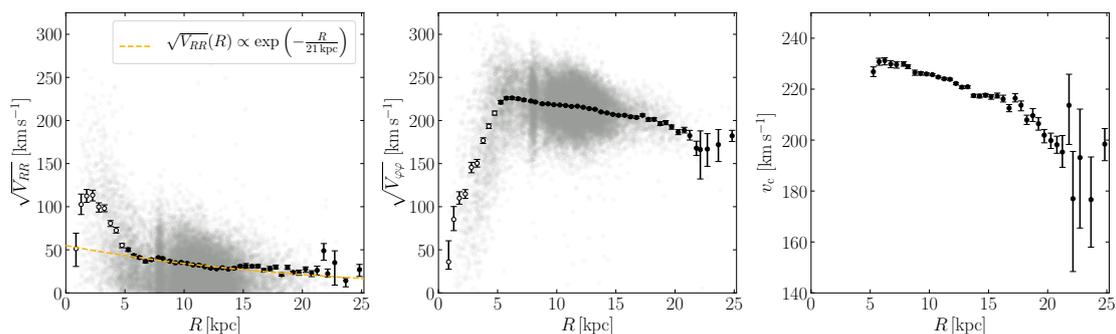
Figure 14.1   Radial profiles of the components of the velocity tensor $\sqrt{V_{RR}}$ (left panel), $\sqrt{V_{\varphi\varphi}}$ (middle panel), and the circular velocity $v_{\rm c}$ (right panel). The small grey dots show individual stars, whereas the black data points show the ensemble averaged values with uncertainties determined via bootstrapping with 100 samples. We exclude all stars within a distance of $R < 5$ kpc from the Galactic center (open data points) from our analysis due to the presence of the Galactic bar.

with a scale length of $R_{\rm exp} = 3$ kpc, which is in good agreement with previous studies (Bland-Hawthorn & Gerhard, 2016). However, the chosen value of the scale length is the dominant source of systematic uncertainty on the circular velocity curve and causes errors at the $\sim 2\%$ level, which we discuss in detail in § 14.3.2.

In order to estimate of the radial dependence of the velocity tensor and to infer the Milky Way's circular velocity, we calculate the radial profiles of the velocity tensor components by averaging all stars within annuli with a minimum width of $\Delta R = 0.5$ kpc or a minimum number of three stars per bin (Fig. 14.1). We then obtain the estimate $\sqrt{V_{RR}}(R)$ from the data itself. We subsequently model this dependency with an exponential function and find a scale length of $R'_{\rm exp} \approx 21$ kpc beyond the inner 5 kpc (yellow dashed curve in the left panel of Fig. 14.1).

Note that within a distance of $R \lesssim 5$ kpc we observe more complex, non-axisymmetric dynamics due to the presence of the bar in the Milky Way. The modeling of the orbits of stars within this region will require to account for deviations from an axisymmetric gravitational potential. However, this is beyond the scope of this paper and hence we exclude this region from our analysis. We can now calculate $v_{\rm c}(R)$ via Eqn. (14.2), which is shown in the right panel of Fig. 14.1.

The maps of the Milky Way colored by the different components of the velocity tensor, as well as the resulting circular velocity are shown in Fig. 14.2.
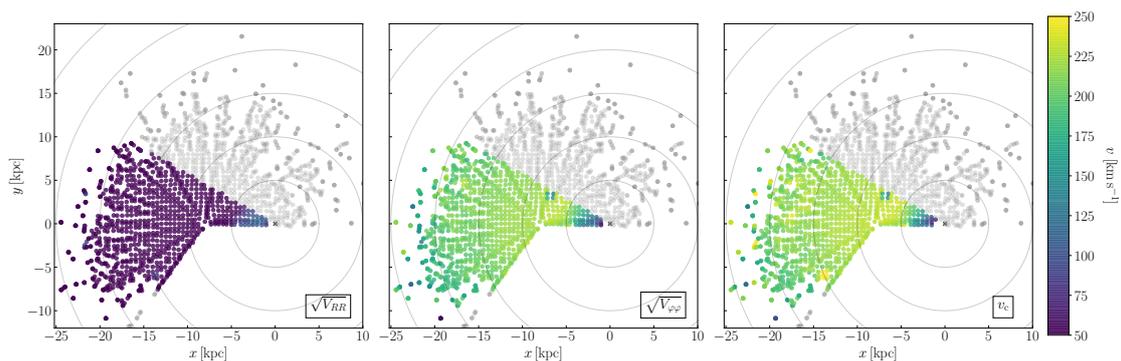
Figure 14.2 Maps of the Milky Way colored by the components of the velocity tensor $\sqrt{V_{RR}}$ (left panel), $\sqrt{V_{\varphi\varphi}}$ (middle panel), and v_c (right panel). Each dot represents an average of the ensemble of stars located within $1\,\text{kpc}^2$ in the $x-$ and $y-$direction. Stars in the grey region are not taken into account for our analysis of the circular velocity curve.

14.3 Results

14.3.1 Measurement of the Circular Velocity Curve

The resulting curve of the Milky Way's circular velocity $v_c(R)$ is shown in Fig. 14.3, while the individual measurements are listed in Table 14.1. It covers a large radial extent between $5 \lesssim R \lesssim 25$ kpc and is very precisely measured, with average uncertainties of $\sigma_{v_c} \lesssim 3\,\text{km}\,\text{s}^{-1}$ determined via bootstrapping from 100 samples. The circular velocity curve shows a gentle but significant decline with increasing radius and can be well approximated by a linear function:

$$v_c(R) = (229.0 \pm 0.2)\,\text{km}\,\text{s}^{-1} - (1.7 \pm 0.1)\,\text{km}\,\text{s}^{-1}\,\text{kpc}^{-1} \cdot (R - R_\odot), \qquad (14.5)$$

which implies a circular velocity with a formal uncertainty of the fit of $v_c(R_\odot) = (229.0 \pm 0.2)\,\text{km}\,\text{s}^{-1}$ at the position of the Sun with a derivative of $(-1.7 \pm 0.1)\,\text{km}\,\text{s}^{-1}$ kpc^{-1}, indicating a constantly, gently declining circular velocity curve.

14.3.2 Systematic Uncertainties

Various systematic uncertainties influence our measurement of the circular velocity curve. In Fig. 14.4 we illustrate the relative systematic differences in the circular velocity $\Delta v_\text{sys}/v_c$ as a function of the Galactocentric radius.

As already mentioned, the largest contribution to the systematic uncertainty of the circular velocity at the location of Sun stems from the unknown density profile of the tracer population. Our fiducial model assumes an exponential function for this density profile with an exponential scale length of $R_\text{exp} = 3\,\text{kpc}$. Varying this

scale length, i.e. $\Delta R_{\mathrm{exp}} = \pm 1\,\mathrm{kpc}$, causes systematic uncertainties at the $\sim 2\%$ level. We also tested the systematic uncertainties arising from the chosen functional form for this density profile. If we instead apply a density profile following a power law, for which we chose an index of $\alpha = -2.7$, which has the same slope as our fiducial exponential function at the location of the Sun, we obtain a systematic error that increases with distance from the Sun to $2-3\%$ for $R \lesssim 20\,\mathrm{kpc}$.

For calculating the circular velocity curve we neglected the term $\frac{\partial v \langle v_R v_z \rangle}{\partial z}$ in the Jeans equation, Eqn. (14.1), which also adds a systematic uncertainty. We estimated this uncertainty to be at the 1% level out to $R \sim 18\,\mathrm{kpc}$, but this term might have larger systematic effects at larger Galactocentric radii.

In order to estimate additional systematic uncertainties on the circular velocity curve arising from our data sample, we split the region, i.e. the $60°$ wedge, within which we consider stars for the analysis, into two disjoint smaller wedges of $30°$ each, and perform the same analysis with both sets of stars. We estimate the systematic uncertainties on the circular velocity by the difference between the resulting fit parameters from the two disjoint data sets. For the circular velocity at the location of the sun we determine a systematic error of $\sigma_{v_c(R_\odot)} \approx 2.02\,\mathrm{km\,s^{-1}}$, which corresponds to uncertainties at the $\lesssim 1\%$ level, whereas the systematic error on the slope of the rotation curve is approximately $0.46\,\mathrm{km\,s^{-1}\,kpc^{-1}}$, which adds uncertainties at the $\sim 27\%$ level.

Further minor systematic uncertainties on the circular velocity at the $\lesssim 1\%$ level stem from the uncertainty on the Galactocentric distance of the Sun, i.e. $\sigma_{R_\odot} = 31\,\mathrm{pc}$ (Gravity Collaboration et al., 2018), and the uncertainties in the proper motion of Sgr A*, i.e. $\sigma_{\mu_l} = 0.026\,\mathrm{mas\,yr^{-1}}$ and $\sigma_{\mu_b} = 0.019\,\mathrm{mas\,yr^{-1}}$ (Reid & Brunthaler, 2004).

All systematic uncertainties taken together affect our measurement of the circular velocity at the $2-5\%$ level out to $R \sim 20\,\mathrm{kpc}$. At Galactocentric distances beyond $20\,\mathrm{kpc}$ the systematic error rises, since our data set only contains relatively few stars at these distances and is thus not very constraining. At the location of the Sun we estimate a systematic uncertainty of $\lesssim 3\%$, which is marked in Fig. 14.3.

We would like to point out that the systematic uncertainties cannot eliminate the gentle decline in the circular velocity curve that we find. In order to obtain a flat circular velocity curve, the exponential scale length of the density profile of the tracer population would has to be very small, i.e. $R_{\mathrm{exp}} < 1$ kpc. Such small scale lengths, however, would imply that almost no stars would be located at Galactocentric distances of $R \sim 20$ kpc and beyond, which is clearly not the case.

Table 14.1   Measurements of the circular velocity of the Milky Way.

$R\,[\mathrm{kpc}]$	$v_{\mathrm{c}}\,[\mathrm{km\,s^{-1}}]$	$\sigma_{v_{\mathrm{c}}}^{-}\,[\mathrm{km\,s^{-1}}]$	$\sigma_{v_{\mathrm{c}}}^{+}\,[\mathrm{km\,s^{-1}}]$
5.27	226.83	1.91	1.90
5.74	230.80	1.43	1.35
6.23	231.20	1.70	1.10
6.73	229.88	1.44	1.32
7.22	229.61	1.37	1.11
7.82	229.91	0.92	0.88
8.19	228.86	0.80	0.67
8.78	226.50	1.07	0.95
9.27	226.20	0.72	0.62
9.76	225.94	0.42	0.52
10.26	225.68	0.44	0.40
10.75	224.73	0.38	0.41
11.25	224.02	0.33	0.54
11.75	223.86	0.40	0.39
12.25	222.23	0.51	0.37
12.74	220.77	0.54	0.46
13.23	220.92	0.57	0.40
13.74	217.47	0.64	0.51
14.24	217.31	0.77	0.66
14.74	217.60	0.65	0.68
15.22	217.07	1.06	0.80
15.74	217.38	0.84	1.07
16.24	216.14	1.20	1.48
16.74	212.52	1.39	1.43
17.25	216.41	1.44	1.85
17.75	213.70	2.22	1.65
18.24	207.89	1.76	1.88
18.74	209.60	2.31	2.77
19.22	206.45	2.54	2.36
19.71	201.91	2.99	2.26
20.27	199.84	3.15	2.89
20.78	198.14	3.33	3.37
21.24	195.30	5.99	6.50
21.80	213.67	15.38	12.18
22.14	176.97	28.58	18.57

22.73	193.11	27.64	19.05
23.66	176.63	18.67	16.74
24.82	198.42	6.50	6.12

Notes. Columns show the Galactocentric radius, the circular velocity, and its negative and positive errorbars.

14.3.3 Velocity Contribution from the Dark Matter Halo

We now proceed to explore in a simple manner how the inferred $v_c(R)$ reflects contributions from different mass components. Due to the linearity of Poisson's equation we can approximate the Milky Way's gravitational potential as a sum of potentials evoked by its individual components i, i.e. the bulge, thin and thick disk, and the halo, hence

$$\Phi = \sum_i \Phi_i \Rightarrow v_c^2 = \sum_i v_{c,i}^2. \tag{14.6}$$

In this initial analysis, we assume that the stellar components are very well determined, and fit only the velocity contribution of the dark matter halo, for which we assume a gravitational potential approximated by an NFW-profile (see Eqn. (2.2); Navarro et al., 1997). For the gravitational potentials of the thin and thick disk we assume Miyamoto-Nagai profiles (Miyamoto & Nagai, 1975), and for the bulge we assume a spherical Plummer potential (Plummer, 1911), while adapting the parameter values of the enclosed mass, the scale length, and the scale height from Pouliasis et al. (2017, model I).

We apply the MCMC affine invariant sampler emcee (Foreman-Mackey et al., 2013), to fit the velocity contribution from the dark matter halo, assuming flat priors for the virial mass $M_{\rm vir} \in [10^{10}, 10^{15}] M_\odot$ and the concentration parameter $c \in [0, 50]$ of the NFW-profile. We adopt the mean of the posterior probability distribution as the best estimate and obtain $M_{\rm vir} = (7.25 \pm 0.25) \cdot 10^{11} M_\odot$ and $c = 12.8 \pm 0.3$, corresponding to a virial radius of $R_{\rm vir} = 189.3 \pm 2.2$ kpc, a scale radius of $R_s = 14.8 \pm 0.4$ kpc, and a characteristic density of $\rho_0 = (1.06 \pm 0.09) \cdot 10^7 M_\odot \, {\rm kpc}^{-3}$. At the location of the Sun the implied dark matter density is $\rho_{\rm dm}(R_\odot) = 0.30 \pm 0.03 \, {\rm GeV \, cm}^{-3}$, presuming the halo is spherical. Our estimate of the velocity contribution of the dark matter halo also reveals that the dynamics of the Milky Way are becoming dominated by dark matter beyond $R \gtrsim 14$ kpc, i.e. the largest contribution to the circular velocity curve beyond this radius originates from the dark matter halo, whereas the inner part is dominated by the stellar components.
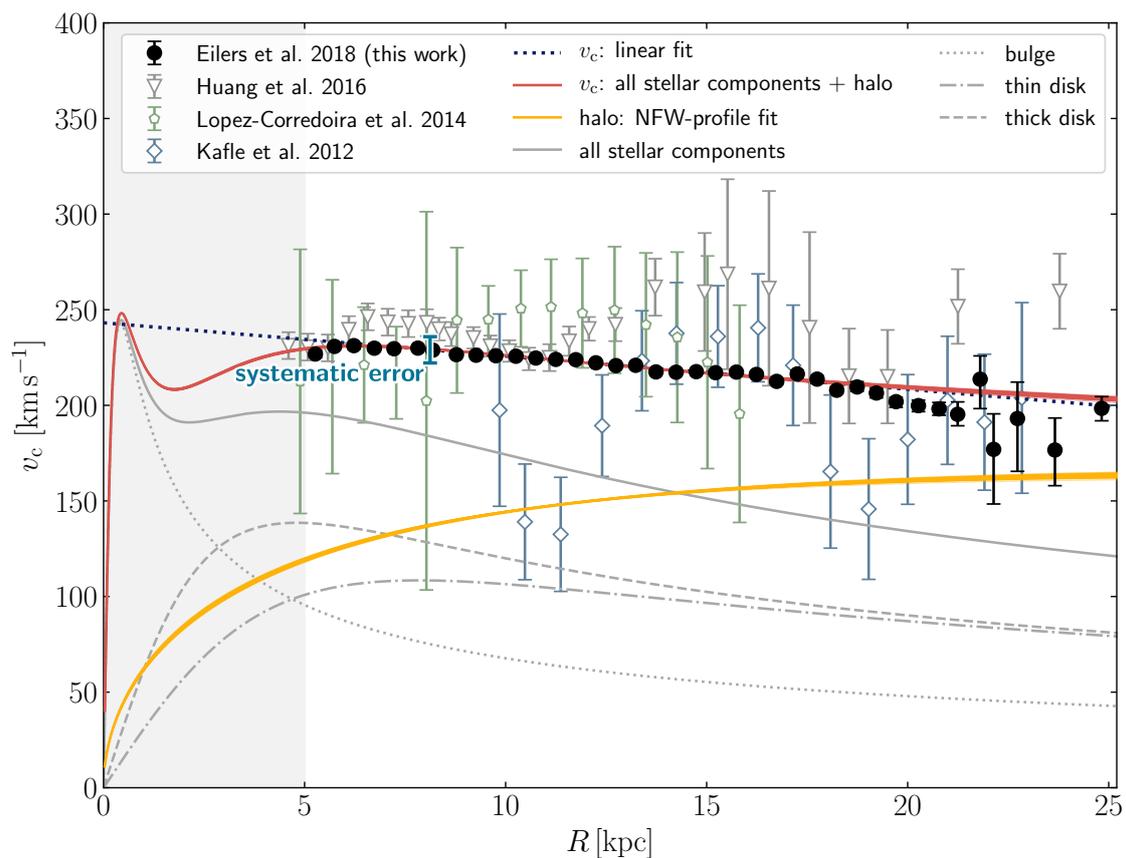
Figure 14.3 The new measurements of the circular velocity curve of the Milky Way are shown as the black data points. The errorbars are estimated via boot-strapping and do not include any systematic uncertainties. We note the systematic error at the location of the Sun, which influences our results at the $\lesssim 3\%$ level (see § 14.3.2). The blue dotted curve shows a linear fit to our data, see Eqn. (14.5), whereas the red curves show 100 random draws from the posterior distribution of the fit parameters to the circular velocity modeled as a sum of stellar components, i.e. bulge, thin and thick disk (grey curves), and a dark matter halo estimated by an NFW-profile (yellow curves, also showing 100 random draws from the posterior). The measurements of various other studies of the circular velocity are shown as colored data points. The light grey shaded region marks the region, where dynamics are strongly influenced by the Milky Way's bar.

14.4 Summary

In this chapter we presented new measurements of the Milky Way's circular velocity curve between $5 \leq R \leq 25$ kpc. *Gaia*'s precise 6-dimensional phase space information for a large sample of tracer stars, together with the new measurements of the Galactocentric rest-frame (Gravity Collaboration et al., 2018), have enabled the most precise measurements of the circular velocity to date. Combining our equatorial circular velocity curve measurement with the kinematics of the Galactic halo
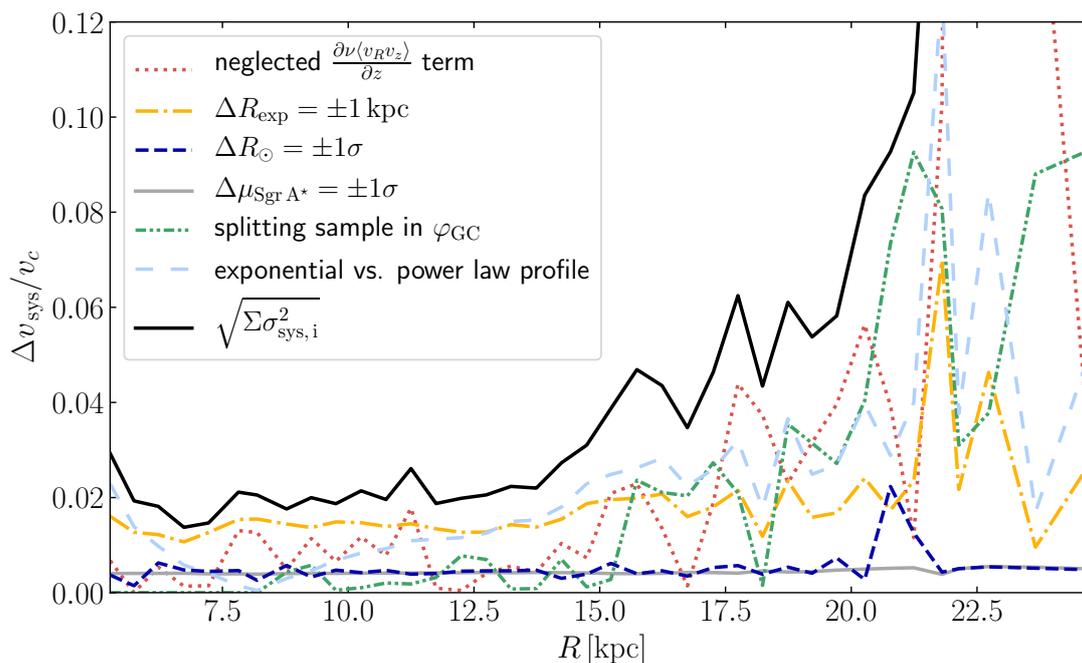
Figure 14.4 Summary of potential systematic uncertainties in the circular velocity curve. We plot the relative deviation in the circular velocity $\Delta v_{\rm sys}$ to our fiducial circular velocity curve v_c, dependent on the Galactocentric radius. We estimate systematic uncertainties arising from the neglected term in Eqn. (14.1) (red dotted curve), from varying the exponential scale length of the density profile of the tracer population (yellow dash-dotted curve), and from changing this density profile from an exponential function to a power law (light blue dashed curve) with an index $\alpha = -2.7$, which has the same slope at the location of the Sun as our fiducial exponential model. Very moderate systematic uncertainties at the $\lesssim 1\%$ level arise from splitting the data sample into two distinct wedges (green dash-dotted curve, see § 14.3.2), varying the distance from the Sun to the Galactic center (blue dashed curve), or uncertainties in the proper motion of Sgr A* (grey curve). All systematic uncertainties added up (black curve) result in uncertainties in v_c at the $\sim 2-5\%$ level out to $R \sim 20\,{\rm kpc}$.

from stellar streams will enable new constraints of the 3-dimensional profile and flattening of the Milky Way's halo (e.g. Xue et al., 2015).

We derive the circular velocity based on a Jeans model under the assumption of an axisymmetric gravitational potential. We used $\gtrsim 23,000$ stars at the upper red giant branch as a tracer population, which are well suited because they are frequent and very luminous, and hence observable over large Galactic distances. We select all stars close to the Galactic plane, because the Milky Way's disk is a cold and precise tracer of kinematic structure. We make use of the fact that red giant stars – similar to red clump stars – are standardizable candles, which allowed us

to obtain precise spectrophotometric distance estimates from a linear data-driven model (see chapter 13).

The dynamics of the inner 5 kpc of the Milky Way disk are strongly influenced by the presence of the Milky Way bar and thus need to be modeled in a non-axisymmetric potential which is beyond the scope of this paper. Beyond $R \gtrsim 5$ kpc, Jeans modelling yields a precise and robust estimate of the circular velocity curve, which shows a gently declining slope of $(-1.7 \pm 0.1)\,\mathrm{km\,s^{-1}\,kpc^{-1}}$, with a systematic uncertainty of $0.46\,\mathrm{km\,s^{-1}\,kpc^{-1}}$. Our result is in reasonably good agreement with another recent analysis of the Milky Way's circular velocity curve based on classical cepheids by Mroz et al. (2018), who find a slope of $(-1.41 \pm 0.21)\,\mathrm{km\,s^{-1}\,kpc^{-1}}$. However, our derivative is significantly less shallow than previous studies by Bovy et al. (2012), Bovy & Rix (2013) or Reid et al. (2014) suggest, who estimate a slope that is consistent with a flat circular velocity curve, which is excluded by our estimate with $> 3\sigma$ significance.

A declining circular velocity curve has not been observed in many other disk galaxies in the local universe, which rather show a flat or even increasing circular velocity curve (e.g. Rubin et al., 1980; Sofue et al., 1999). Galaxies with declining circular velocity curves have yet only been reported at higher redshift. For instance, Genzel et al. (2017) studied six massive star-forming galaxies at $z \approx 2$ and found declining circular velocities curves, claiming that these galaxies are baryon-dominated and their dark matter content smaller than in disk galaxies in the local universe (see also Lang et al., 2017). They argue that the observations suggest that baryons in the early universe during the peak epoch of star formation efficiently condense at the centers of dark matter halos when gas fractions are higher and dark matter is less concentrated.

Our estimate of the circular velocity at the Sun's position of $v_\mathrm{c}(R_\odot) = (229.0 \pm 0.2)\,\mathrm{km\,s^{-1}}$ is very precisely determined, with systematic uncertainties at the $\sim 2 - 3\%$ level. Despite the formally very small uncertainties in our circular velocity curve, the systematic uncertainties introduced by various factors, such as the choice of the exponential scale length of the tracer population for instance, see Eqn. (14.4), the assumed axisymmetry of the gravitational potential, or the *APOGEE* selection function could bias our results. These systematic uncertainties might increase at large Galactocentric distances, i.e. $R \gtrsim 20$ kpc, where our data is not very constraining. Our estimate is in good agreement with the estimate by Koposov et al. (2010), i.e. $v_\mathrm{c}(R_\odot) = 221^{+16}_{-20}\,\mathrm{km\,s^{-1}}$, and Wegg et al. (2018), who estimated $v_\mathrm{c}(R_\odot) = (222 \pm 6)\,\mathrm{km\,s^{-1}}$ from modeling RR Lyrae stars with *Gaia*. However, it is significantly lower than the measurement by Reid et al. (2014), i.e.

$v_{\rm c}(R_\odot) = (240 \pm 8)\,{\rm km\,s^{-1}}$, but higher than the estimate of $v_{\rm c}(R_\odot) = (218 \pm 6)\,{\rm km\,s^{-1}}$ by Bovy et al. (2012). We also calculated the circular velocity from combining the result for the Sun's Galactocentric distance by Gravity Collaboration et al. (2018) with the recently published Oort's constants by Bovy (2017), which he derived from main-sequence stars from the *Gaia* DR1 Tycho-Gaia Astrometric Solution (TGAS), and obtain $v_{\rm c}(R_\odot) = (220.9 \pm 4.7)\,{\rm km\,s^{-1}}$, which is about 2σ lower than our estimate.

The very smooth shape of our circular velocity curve is in broad agreement with previous work. However, we do not see any evidence for a dip in $v_{\rm c}$ at $R \approx 9$ kpc or $R \approx 11$ kpc that was previously claimed by other studies (Sofue et al., 2009; Kafle et al., 2012; Huang et al., 2016; McGaugh, 2018) and had been interpreted as potential signatures of spiral arms. We do see a mild but significant deviation from the straightly declining circular velocity curve at $R \approx 19 - 21$ kpc of $\Delta v \approx 15\,{\rm km\,s^{-1}}$. Differences in the estimation of $v_{\rm c}(R)$ between our work and previous studies could potentially be due to different tracer populations that have been used for the analyses. Kafle et al. (2012) for instance used blue horizontal branch stars in the halo, i.e. $|z| > 4$ kpc, whereas our analysis is focused on stars within the Milky Way disk.

Combined with other work for the mass distribution of the stellar components of the Milky Way, we estimated the NFW-profile of the dark matter halo, while keeping contributions from the baryonic components fixed. Our estimate of the virial mass, $M_{\rm vir} = (7.25 \pm 0.25) \cdot 10^{11}\,M_\odot$, is significantly lower than what several previous studies suggest. The recently published analysis by Watkins et al. (2018) based on the kinematics of halo globular clusters determined by *Gaia* DR2 data, measures a virial mass of $M_{\rm vir} = 1.41^{+1.99}_{-0.52}\,M_\odot$ that is roughly twice as high than our value, although the measurements have large uncertainties (see also Piffl et al., 2014; Küpper et al., 2015; Bland-Hawthorn & Gerhard, 2016; Huang et al., 2016). However, virial mass measurements that agree well with our estimate are presented by Bovy et al. (2012), as well as Eadie & Harris (2016) and Eadie & Jurić (2018), whose analyses are similar to the one from Watkins et al. (2018) and also based on kinematic data of globular clusters.

Our estimated local dark matter density is in good agreement within the uncertainties with the estimated values by Huang et al. (2016), who measured $\rho_{\odot,{\rm dm}} = 0.32 \pm 0.02\,{\rm GeV\,cm^{-3}}$, and by Zhang et al. (2013), who estimated $\rho_{\odot,{\rm dm}} = 0.25 \pm 0.09\,{\rm GeV\,cm^{-3}}$, although significantly lower than $\rho_{\odot,{\rm dm}} = 0.542 \pm 0.042\,{\rm GeV\,cm^{-3}}$ estimated by Bienaymé et al. (2014). Additionally, our new measurement of the circular velocity curve covering large Galactocentric distances, enabled us to determine the Galactic radius beyond which the mass of the Milky Way is dominated by

dark matter, which we estimate to be $R \gtrsim 14$ kpc. This value is in vague agreement with the result by Bovy & Rix (2013), who found that the Milky Way's circular velocity is dominated by the stellar components at $R < 10$ kpc. A similar study by Portail et al. (2017) obtains a distance of $R \approx 8$ kpc that marks the transition from the baryonic to dark matter dominated regime. Obviously, our inferred properties of the Milky Way's dark matter halo are dependent on the measurements of the baryonic mass components, which we keep fixed in our analysis. Indeed there is some disagreement in the literature about the exact stellar masses, and hence if the measurements from Pouliasis et al. (2017), which we use in our analysis, are overestimated as suggested by Portail et al. (2017), our results would be underestimating the influence of the dark matter halo.

Part IV

Discussion and Conclusions

Quasars in the Early Universe: Summary & Outlook

In this chapter we summarize briefly the main results of part II of this thesis (§ 15.1) and outline several projects we are planing to conduct in the future based on our results (§ 15.2).

15.1   Summary

The fundamental data set for this thesis is a new sample of 34 high redshift quasar spectra at $5.77 \leq z_{\mathrm{em}} \leq 6.54$. The spectra have all been observed with ESI on the Keck telescopes at optical wavelengths (4000Å – 10000Å) with a resolution of $R \approx 4000 - 5400$, and include data that we collected along with observations from the Keck Observatory Archive. All exposures from different observing runs have been consistently reduced and co-added, resulting in a very rich and homogeneous data set, comprising a total of ~ 180 hours of telescope time. The final spectra have been publicly released (see § 8.4).

In this thesis we conduct various analyses based on this data set. The main results are:

- The redshift evolution of the IGM opacity constrains the timing and morphology of the Epoch of Reionization, and the thermal state of the post-reionization IGM, both of which influence early galaxy formation. By analyzing the IGM opacity along the quasar sightlines in our data sample in Chapter 9, we find an increased scatter in the IGM opacity of the Lyα forest down to $z \sim 5$, providing evidence for a more extended reionization epoch than previously thought.

We compare our measurements to a large-volume hydrodynamical simulation with a uniform UVB and find that the spread in observed τ_{eff} cannot be explained by fluctuations in the underlying density field alone, supporting an inhomogeneous reionization scenario, which is consistent with previous analyses. A comparison to semi-numerical simulations with additional fluctuations from a fluctuating UVB or temperature field can both account for the observed scatter in our measurements.

- A subsequent analysis of the opacity within the Lyβ forest in Chapter 10 (Eilers et al., to be submitted) suggests much stronger fluctuations in the UVB of current reionization models than previously assumed, once we forward-model the spectral noise of the data. At $z \sim 6$ our analysis also reveals a mismatch in the mean opacity of high-density regions in the IGM between observations and state-of-the-art hydrodynamical simulations, possibly suggesting an inversion of the thermal state of the IGM. Further modeling will be required to explain the observed opacity both in the Lyα and the Lyβ forest, which will lead to new insights regarding the morphology, timing, and evolution of the reionization process.

- In Chapter 11 we analyze the sizes of the proximity zones for our sample of high redshift quasar spectra. Previous studies had found a strong evolution of proximity zone sizes with redshift, and it was argued that this reflected the rapid evolution of the IGM neutral fraction during the Epoch of Reionization. Contrary to these results we find a much shallower redshift evolution, which is however consistent with the evolution predicted by our radiative transfer simulations, irrespective of assumptions about the ionization state of the IGM.

This result can be explained, considering that the size of the proximity zone ends at a distance corresponding to a flux transmission level of 10% according to the definition of quasar proximity zones (Fan et al., 2006c) and, in a highly ionized IGM, this distance is reached at a location where the ionization rate of the quasar dominates the ionization rate of the UVB by at least an order of magnitude. As such, the size of the quasar proximity zone is essentially independent of the UVB and neutral gas fraction. Assuming a neutral ambient IGM the observed proximity zone sizes R_p track the growth of the ionized H II region R_{ion}, but then cease to grow when the distance with a 10% flux transmission level is reached. For a highly plausible quasar age of $t_Q \sim 10^{7.5}$ yr, R_p is thus independent of the ionization state of the IGM

also for a mostly neutral IGM. Thus contrary to previous claims, both the observed shallow redshift evolution and the results from our simulations imply that the redshift evolution of proximity zone size R_p does not significantly constrain the evolution of the neutral gas fraction during the Epoch of Reionization.

- We show for the first time how the sizes of the proximity zones can give independent constraints on the quasars' lifetimes with unprecedented precision. These proximity zones are sensitive to the lifetime of the quasars, because the intergalactic gas has a finite response time to the quasars' radiation. A schematic description of the principle of this novel method is shown in Fig. 15.1. In order to obtain constraints on the lifetimes from measurements of quasar proximity zones that go beyond toy models we compare the observations to simulated quasar spectra at various lifetimes. We make use of the cosmological hydrodynamical simulation Nyx (Almgren et al., 2013), and post-processed the sight lines through the simulation box with a 1-dimensional radiative-transfer code (Davies et al., 2016) to simulate the effects of the quasars' radiation. A comparison of the measurements to the simulated sizes of proximity zones allows us to determine precise lifetimes for individual quasars.

  By analyzing the proximity zones of all quasars in our data set, we discover an unexpected population of very young quasars ($t_Q \lesssim 10^4 - 10^5$ yr) that exhibit exceptionally small proximity zones. These 10% of quasars pose significant challenges on current black hole formation theories, because their lifetime is several orders of magnitude too short to grow their central SMBHs when assuming a growth rate at which the quasars shine at the Eddington luminosity, and may require some new black hole formation mechanism.

- We conduct the first spectroscopic study of the youngest quasar known to date SDSS J1335+3533 in Chapter 12, with a lifetime of only $t_Q = 10^{3.0\pm0.8}$ yr. With our new high S/N data we are able to rule out any premature truncation of the proximity zone due to a self-shielding absorption system. By means of the Mg II line we measure the mass of the SMBH of the quasar and find a billion solar mass black hole in its center, which is in strong tension with the estimated quasar age.

  The spectrum of SDSS J1335+3533, shows no evidence for any spectral anomalies associated with its young age, with the potential exception that it is a WLQ. In order to make further progress, we will need to identify larger sam-
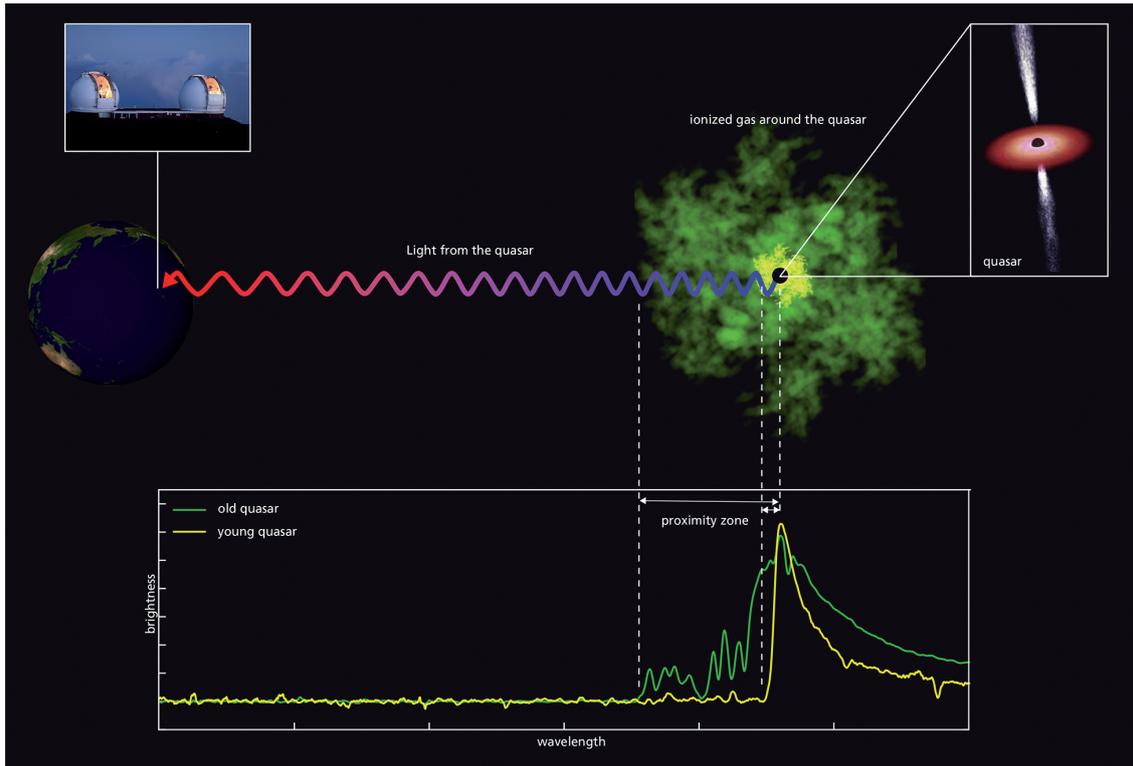
Figure 15.1 Schematic description of the new method to determine the lifetime of quasars based on the sizes of their proximity zones (Figure Credit: A.-C. Eilers and J. Neidel).

ples of young quasars, which might elucidate the connection between WLQs and quasar youth.

We discuss several scenarios that could potentially explain the rapid black hole growth and design observations with an IFU that will distinguish between different formation scenarios by measuring the surface brightness profile of nebular emission lines within the ENLR. We estimate that with the upcoming capabilities on JWST we can measure these surface brightness profiles with only modest integration times.

15.2 Future Directions

One of the main results of this thesis, the novel and innovative approach to determine quasar lifetimes, promises new insights into early black hole formation. It provides the unique opportunity to estimate lifetimes of individual quasars in the early universe and hence explore the evolution of spectral and environmental features with quasar age. We hope that the results of these studies will provide answers to the long-standing question of how SMBHs form, how quasar activity is

triggered, and how galaxies and quasars co-evolve, which is of significant importance for galaxy formation models at all redshifts. With the aim to answer these question we intend to pursue several ideas in the future as outlined below.

15.2.1 Distribution of Lifetimes for the Quasar Population at Large

Our new method to determine quasar lifetimes from the sizes of their ionized proximity zones offers the unique possibility to measure the distribution of lifetimes for the whole quasar population. This distribution of lifetimes, which we ultimately intend to determine for a luminosity-complete sample of high redshift quasars, will enable unparalleled observational constraints on early growth phases of SMBHs.

However, in order to place tight constraints on lifetimes, precise estimates of the quasars' systemic redshifts are necessary, since these constitute the largest source of uncertainty in proximity zone measurements. To this end, in Atacama Large Millimetre Array (ALMA) cycle 5 we had a successful A-ranked program (PI: Eilers), as well as a parallel approved program with the Northern Extended Millimeter Array (NOEMA) on Plateau de Bure (PI: Eilers), to follow up ten new young quasar candidates at $z \gtrsim 6$ with the goal to detect the $158\,\mu$m fine-structure [CII] or the CO (6-5) 3 mm emission line arising from the molecular gas reservoir of the quasars' host galaxies and hence providing the most precise redshift estimates. Our preliminary results are shown in Fig. 15.2. We aim to precisely determine the quasars' systemic redshifts from these sub-mm observations and estimate the quasars' lifetimes.

In order to find the best targets for the ALMA and NOEMA programs, we analyzed a sample of ~ 350 discovery quasar spectra at $z \gtrsim 5.8$. Assuming a fraction of 10% young quasars in the quasar population at large, which is consistent with our detection rate, we expect to identify $3 - 4$ new young quasars with this data.

15.2.2 Multi-Wavelength Characterization of Young Quasars

We aim to undertake a multi-wavelength study, from sub-mm wavelengths to X-ray observations, in order to fully characterize and understand the properties of young quasars. In particular, our recently obtained sub-mm observations will potentially reveal traces of an active merger or any companion galaxies in the vicinity of the quasars, whose recent encounter or interaction might have triggered the quasar activity of the young objects. We will propose to obtain X-ray observations, which trace the innermost regions of the accretion disk and allow direct constraints on
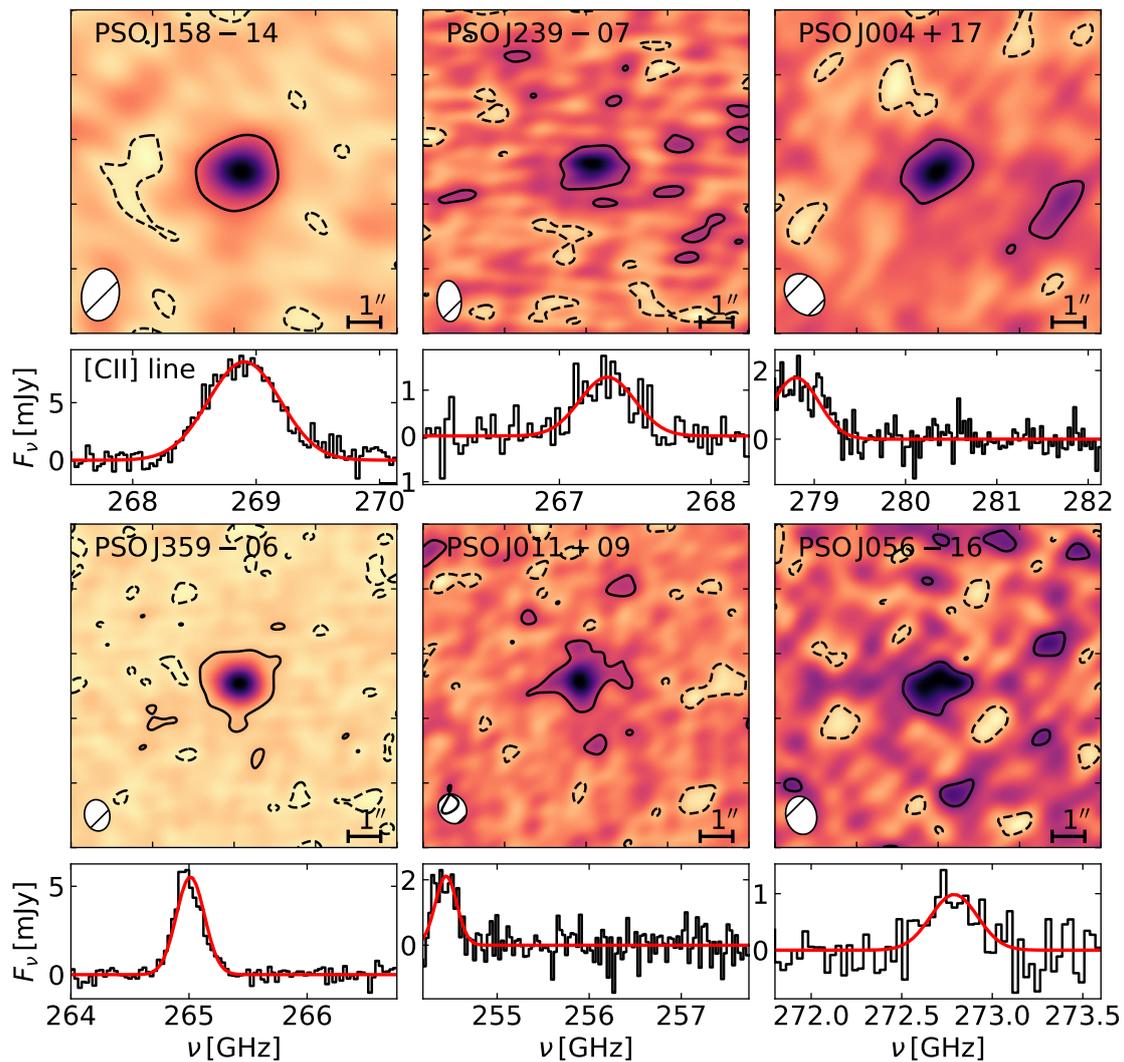
Figure 15.2 ALMA observations of six potentially new young quasars. Detection of the narrow [CII] emission line for six out of ten new young quasar candidates, in order to determine the systemic redshifts of their host galaxies, which are necessary to obtain precise estimate of the quasars' lifetimes (Eilers et al. in prep.).

the mass accretion rate of the quasars (Shemmer et al., 2009, 2010). A more rapid accretion rate than the theoretical limit, i.e. super-critical mass accretion rates, could explain the growth of SMBHs in quasars over a very short lifetime.

We have recently been leading a program to obtain deep optical and NIR data from the X-Shooter spectrograph on the VLT (PI: Eilers) as well as from DEIMOS on the Keck telescopes (PI: Hennawi), which will allow us to search for spectral as well as environmental signatures that might distinguish the young quasars from the quasar population at large. In our spectroscopic study of one of the young quasars we proposed that weak emission lines could be connected to the youth of
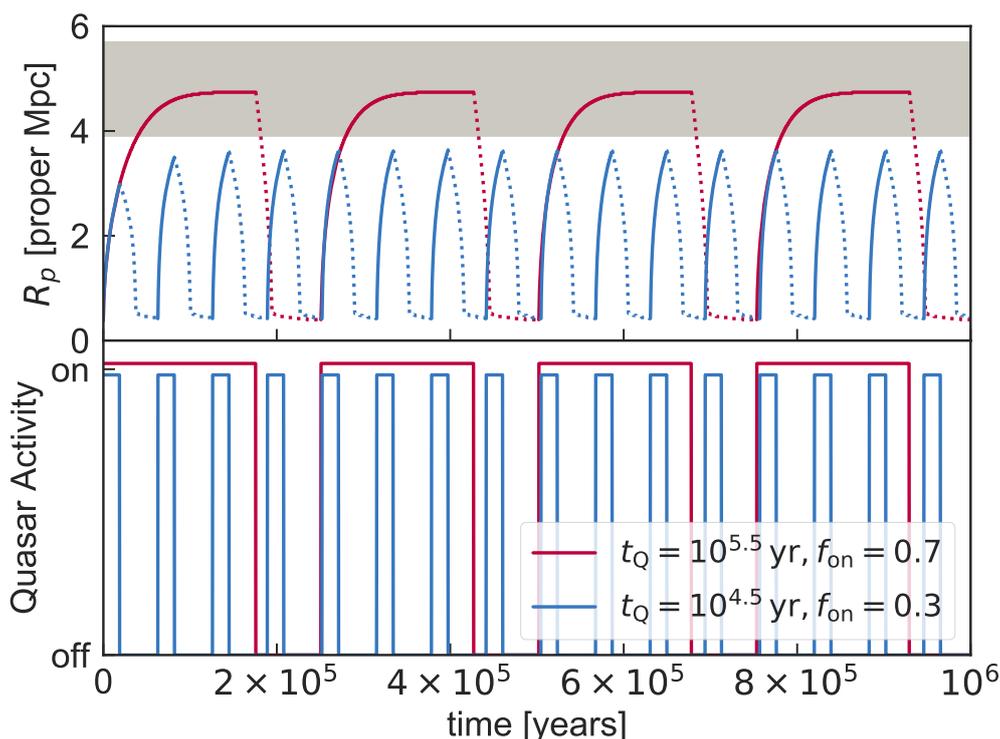
Figure 15.3   Semi-analytic model of a flickering quasar light curve. The resulting proximity zones (*top*) from two different quasar light curves (*bottom*) with varying quasar lifetime and fraction of time that the galaxy spends in the active luminous quasar phase f_{on}. The grey shaded region shows the distribution of observed proximity zones, which clearly excludes the blue quasar light curve. The dotted parts of the curves indicate epochs at which the galaxy is in a quiescent phase, i.e. the quasar is "off", and hence its proximity zone could not be observed (Davies, Hennawi, & Eilers, to be submitted).

a quasar (see Chapter 12), suggesting that the quasar's radiation did not yet have enough time to illuminate its BLR. A confirmation of this hypothesis would also solve the long-standing mystery about the origin of WLQs. This data will also allow us to estimate the masses of the central SMBHs, the quasars' emission line ratios, determine whether strong in– or outflows influence the broad line region of the quasars, and constrain the star formation history of the host galaxies by estimating the relative abundances of metals in the broad line region.

15.2.3   Quasar Light Curves and Quasar Variability

Knowing the distribution of lifetimes for the quasar population at large at high redshift will also allow constraints on the shape of quasar light curves. In particular, we intend to analyze galaxy-quasar co-evolution models in which active luminous quasar and quiescent galaxy phases alternate, and assess which constraints on the

time scales of active and quiescent periods we can obtain from proximity zone measurements. We recently developed a new semi-analytic model for proximity zones based on hydrodynamical simulations post-processed with a 1-dimensional radiative transfer code (Davies, Hennawi, & Eilers to be submitted). This model allows us to simulate the distribution of quasar proximity zones while varying the quasar lifetimes and the fraction of time f_{on} that the galaxy is in an active quasar state. Two example light curves are shown in Fig. 15.3.

We aim to build upon this semi-analytic model and analyze more realistic, stochastic quasar light curves with a distribution of various Eddington ratios, i.e. different mass accretion rates, from radiative transfer simulations (e.g. Novak et al., 2011) or modeled by a damped random walk (e.g. Sartori et al., 2018), and determine the distribution of proximity zone sizes from these models. A comparison of the modeled to the observed distribution of proximity zones and quasar lifetimes will enable new insights into quasar variability on various time scales by constraining the shape of their light curves.

15.2.4 New Constraints on Quasar Lifetimes

We plan to investigate several new approaches to place constraints on quasar lifetimes.

Extent of Radio Jets

A completely unique opportunity to independently constrain a quasar's lifetime is provided by the recently discovered radio-loud quasar $PSO J352 - 15$ at $z \approx 5.84$ (Bañados et al., 2018a; Momjian et al., 2018), which presents the first clear evidence of radio jets at $z \sim 6$ (Fig. 15.4). The extent of these powerful radio jets allows an estimate of the quasar's lifetime after careful consideration of the geometry of its alignment. We intend to analyze this source and constrain its lifetime with two independent approaches based on its geometry and its proximity zone size, making use of new sub-mm data from a recently approved, A-ranked program on NOEMA (PIs: Eilers & Bañados).

Surface Brightness of Lyα Nebulae or the Extended Narrow Line Regions

Additionally, IFU observations of the surface brightness of the Lyα nebulae around quasars or the extended narrow line region traced by near-infrared nebular emission lines (such as Hα, Hβ, or [O III]), will provide new constraints on the quasars' lifetimes, as outlined in § 12.4. In particular, these IFU observations will allow
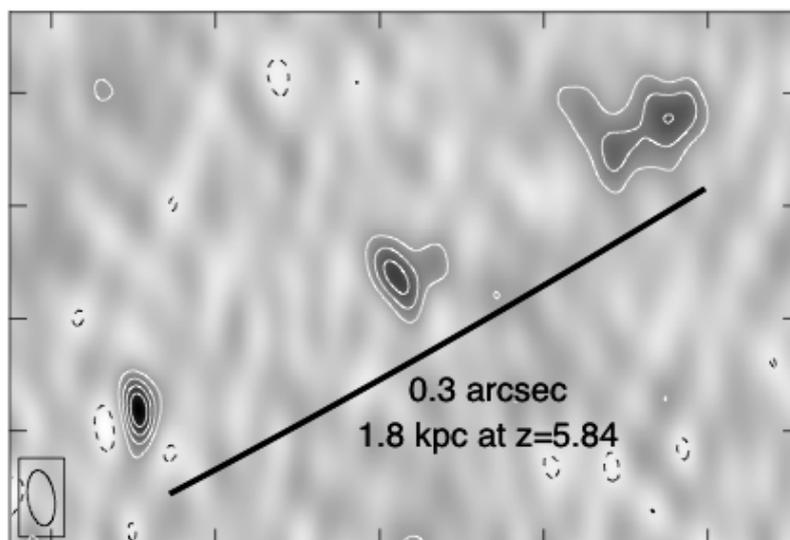
Figure 15.4 Image of the first radio-loud quasar at $z \sim 6$ with clear evidence for radio jets. The object PSO J352−15 is resolved on three distinct regions, resembling a classic Fanaroff-Riley II morphology, with a radio core and two radio lobes. The source is small, from hot spot to hot spot only ~ 1.8 physical kpc, indicating that the quasar must be young. The image is taken with the Very Long Baseline Array (VLBA) at 1.5 GHz and 20 mas resolution (Momjian et al., 2018).

us to distinguish between different scenarios that could be explaining the very small proximity zones of the young quasars. Furthermore, these observations will yield insights into the large-scale environment of the quasars, and determine whether strong in– or outflows influence the quasars' surrounding and potentially the growth of their SMBHs.

We intend to conduct the IFU observations from the ground with the new Large Lenslet Array Magellan Spectrograph (LLAMAS) on the Magellan telescope, which will be commissioned in the summer of 2019, or the Multi Unit Spectroscopic Explorer (MUSE) on the VLT. These observations will be a pathfinder for the upcoming capabilities on JWST, which will allow a deep follow-up of the most exciting young quasars with only very modest integration times (see § 12.4).

Helium Proximity Zones

In addition to proximity zones measured in the H I Lyα forest, we are aiming to also characterize proximity zones in the Lyα forest of He II of quasars at $z \sim 4$, leveraging far-UV spectra from the HST/COS spectrograph. The He II sightlines provide a new approach for measuring intermediate quasar lifetimes. Whereas for H I proximity zones the dependence of R_p on t_Q flattens once ionization equilibrium
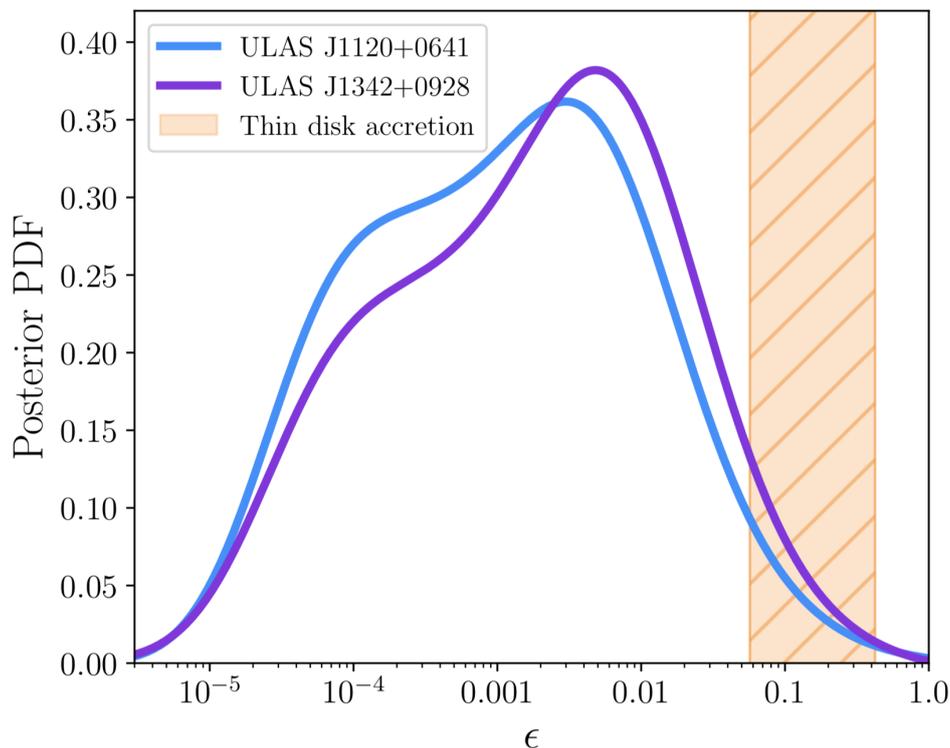
Figure 15.5 Evidence for super-critical accretion rates for two quasars at $z >$ 7. Colored curves show the posterior probability distributions for the radiative efficiency ϵ for the two highest redshift quasars known, suggesting much lower radiative efficiencies, i.e. higher mass accretion rates, than usually assumed in thin disk quasar accretion models (Davies, Hennawi, & Eilers, subm.).

is reached, i.e. at $t_Q > t_{eq} \gtrsim 3 \times 10^4$ (§ 6.3), the sizes of He II proximity zones retain sensitivity up to $t_Q \approx 3 \times 10^7$ yr, due to the higher ionization energy of He II and hence the longer equilibration timescale (Khrykin et al., 2018).

Damping Wing Feature in $z \gtrsim 7$ Quasars

The damping wing feature observed in quasars at very high redshift $z \gtrsim 7$ (Simcoe et al., 2012; Davies et al., 2018c; Bañados et al., 2018b), where the surrounding IGM is still significantly neutral, provides a measurement of the lifetime *integrated* over all active phases and hence a statistical sample of high redshift quasars will provide further insights onto early black hole formation and the lifetimes of quasar.

One potential explanation for the rapid assembly of SMBHs in young quasars, could be a higher mass accretion rate than expected in the standard picture for black hole growth. Building upon our new method to infer quasar lifetimes, we recently determined the lifetimes of the two most distant quasars known to date

at $z > 7$. Detailed modeling of the proximity zones and damping wing at these high redshifts reveals a degeneracy between the quasar lifetime and the neutral gas fraction of the IGM. However, after marginalizing over these parameters, we obtain a constraint on the radiative efficiency of the accretion of the quasars.

We find evidence for much lower radiative efficiencies (see Fig. 15.5), i.e. much higher, super critical accretion rates (Davies, Hennawi, & Eilers, subm.). This result suggests mass accretion rates several orders of magnitude higher than the theoretical limit, which are required in order to explain the formation of the SMBHs.

Structure and Dynamics of the Milky Way: Summary & Outlook

We will summarize here briefly the results of part III of this thesis (§ 16.1), and present a few ideas for the future development of the project (§ 16.2).

16.1 Summary

One major goal of the upcoming *SDSS-V* survey is to enable large global maps of our own galaxy. It targets about 4 million stars in the Milky Way and aims to decipher the chemical and dynamical evolution of the Galaxy. However, in order to achieve this goal precise distance estimates are necessary. While *Gaia* provides great measurements of proper motions, uncertainties in the distances dominate all studies beyond a few kpc distance from the Sun.

Therefore we develop a new data-driven model to determine precise parallaxes of $\gtrsim 45,000$ luminous red giant stars by combining multi-band photometric information from *Gaia*, *2MASS*, and *WISE* with spectroscopy from the *APOGEE* survey and *Gaia* DR2 parallaxes, which we presented in Chapter 13. Our model assumes that red giant stars are dust-correctable and standardizable candles and that the spectroscopic and photometric features are sufficient to predict the logarithmic parallax – and hence the distance modulus – by a purely linear function of the chosen features.

We split our data set randomly in two distinct subsets A and B to optimize our model on one set and apply the optimized coefficient vector to predict parallaxes for the other part of the data set. In a cross-validation test our model predicts spectrophotometric parallaxes with a scatter of $\sim 9-10\%$. These predicted paral-

laxes are more than twice as precise as *Gaia*'s measurements already for distances beyond $\gtrsim 1$ kpc for stars with a magnitude of $G \sim 14$ mag.

This enables precise kinematic maps of our Galaxy out to Galactocentric distances of 25 kpc, well beyond the reach of *Gaia* parallaxes. The catalog of the spectrophotometric parallaxes has been publicly released via the zenodo platform.

Making use of these new spectrophotometric parallaxes, we determine in Chapter 14 the most precise measurement to date of the circular velocity curve $v_c(R)$ of the Milky Way over a wide range of Galactocentric distances. Based on Jeans modeling in an axisymmetric gravitational potential we find that the circular velocity curve is gently but significantly declining with a slope of $(-1.7 \pm 0.1)\,\mathrm{km\,s^{-1}\,kpc^{-1}}$. Declining circular velocity curves have so far only been reported for galaxies at higher redshifts (e.g. Genzel et al., 2017). A flat rotation curve, however, which has been observed in most other galaxies in the local universe, can be excluded with $> 3\sigma$ significance.

At the location of the Sun we measure a velocity of $v_c(R_\odot) = 229.0 \pm 0.2\,\mathrm{km\,s^{-1}}$. A detailed analysis of potential systematic uncertainties, such as the density profile of the tracer population or uncertainties in the distance and proper motion of the Galactic center for example, reveals that these uncertainties influence our result at the $2-3\%$ level.

Decomposing the circular velocity curve into various stellar and dark matter components, we estimate a dark matter density at the location of the Sun of $\rho_{\odot,\mathrm{dm}} = 0.32 \pm 0.02\,\mathrm{GeV\,cm^{-3}}$, which provides constraints for any direct as well as indirect dark matter detection experiments.

16.2 Future Directions

16.2.1 Non-axisymmetric Structure of the Milky Way

Structures of the Milky Way, such as its spiral arms or the Galactic bar cannot be modeled by a simple axisymmetric gravitational potential. Leveraging our precise spectrophotometric parallaxes over a large range of Galactocentric distances we aim to analyze these non-axisymmetric features, which can be revealed by the cross correlation coefficients between radial and tangential velocity components of the stars. The combination of our measurements with data from *APOGEE* South will equip us with a complete kinematic map of the Milky Way disk to perform these studies.
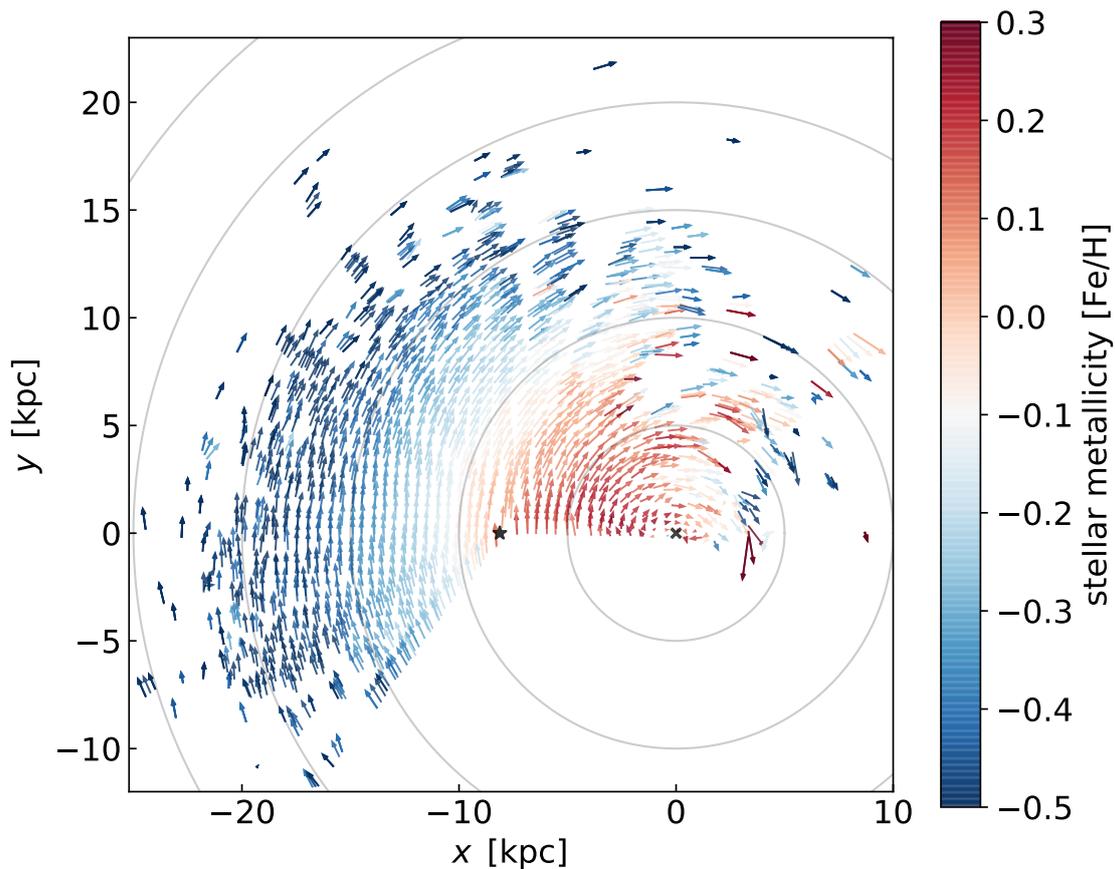
Figure 16.1 Map of the Milky Way color coded by the mean stellar metallicity [Fe/H]. Arrows indicate the mean motion of stars. The Galactic center is marked by the black cross, and the Sun's position is indicated by the black star (Eilers et al., in prep.).

Additionally, we plan to model the sources of the non-axisymmetric components of our Galaxy and analyze whether the spiral pattern arises due to the spiral arms of the Milky Way, potential resonances from the Galactic bar, or whether it could have been invoked by a recent encounter with the Sagittarius dwarf galaxy (e.g. Purcell et al., 2011). Other non-axisymmetric features that could only be revealed in studies covering large Galactic scales would be a potential warp in the outer part of the Milky Way disk, as well as the location, extension, and angle of the bar in the inner part of the disk.

16.2.2   Element Abundance Gradients

The observed stellar abundance gradients in the Galactic disk and its evolution offer the opportunity to test theories of disk chemical evolution and stellar nucleosynthesis. Detailed knowledge about these gradients are crucial for our un-

derstanding of the metal enrichment and star formation history. Combining our new spectrophotometric distances with the 15 different element abundances delivered by the *APOGEE* survey, will enable us to obtain detailed abundance gradients across large Galactocentric distances for various elements. As an example, Fig. 16.1 shows a map of the Galactic disk of the Milky Way color coded by the average stellar metallicity [Fe/H]. We aim to analyze the effects of the Galactic bar on the chemical composition in the inner region of the Milky Way, as well as the effects of radial migration for the evolution of the outer part of the Galactic disk.

Additional Information on the Data Set of Quasar Spectra

Table A.1 Overview of the observation runs for the 34 quasars in our data sample.

object	RA (J2000)	DEC (J2000)	PI	date	exp. time
SDSS J0002+2550	$00^h02^m39\overset{s}{.}39$	$+25°50'34\overset{''}{.}96$	Kakazu	Nov. 2004	5400 s
			Cowie	Aug. 2005	16300 s
SDSS J0005-0006	$00^h05^m52\overset{s}{.}34$	$-00°06'55\overset{''}{.}80$	Becker	Dec. 2002	1200 s
			Sargent	Oct. 2010	15000 s
CFHQS J0050+3445	$00^h55^m02\overset{s}{.}91$	$+34°45'21\overset{''}{.}65$	Willott	Sep. 2008	6250 s
			Sargent	Oct. 2010	6000 s
			Sargent	Oct. 2012	9000 s
SDSS J0100+2802	$01^h00^m13\overset{s}{.}02$	$+28°02'25\overset{''}{.}92$	White	Jan. 2016	3600 s
ULAS J0148+0600	$01^h48^m37\overset{s}{.}64$	$+06°00'20\overset{''}{.}06$	Sargent	Oct. 2010	11200 s
ULAS J0203+0012	$02^h03^m32\overset{s}{.}38$	$+00°12'29\overset{''}{.}27$	Sargent	Oct. 2010	6600 s
CFHQS J0210-0456	$02^h10^m13\overset{s}{.}19$	$-04°56'20\overset{''}{.}90$	Sargent	Oct. 2010	6000 s
PSO J036+03	$02^h26^m01\overset{s}{.}87$	$+03°02'59\overset{''}{.}42$	White	Jan. 2016	11700 s
CFHQS J0227-0605	$02^h27^m43\overset{s}{.}29$	$-06°05'30\overset{''}{.}20$	Willott	Sep. 2008	10540 s
SDSS J0303-0019	$03^h03^m31\overset{s}{.}40$	$-00°19'12\overset{''}{.}90$	Sargent	Oct. 2010	6000 s
SDSS J0353+0104	$03^h53^m49\overset{s}{.}73$	$+01°04'04\overset{''}{.}66$	Sargent	Oct. 2010	13200 s
			Becker	Jan. 2006	3600 s
PSO J060+24	$04^h02^m12\overset{s}{.}69$	$+24°51'24\overset{''}{.}43$	White	Jan. 2016	10800 s
SDSS J0818+1723	$08^h18^m27\overset{s}{.}40$	$+17°22'52\overset{''}{.}01$	Becker	Apr. 2005	2400 s
SDSS J0836+0054	$08^h36^m43\overset{s}{.}86$	$+00°54'53\overset{''}{.}26$	Becker	Mar. 2001	1731 s
			Madau	Feb. 2002	18900 s
			Cowie	Feb. 2002	13540 s

			Djorgovski	Mar. 2002	8923 s
			Kulkarni	Jan. 2003	5400 s
			Cowie	Jan. 2003	3600 s
			Cowie	Jan. 2004	10800 s
			Kakazu	Nov. 2004	10200 s
			Djorgovski	Dec. 2004	7200 s
SDSS J0840+5624	$08^{\mathrm{h}}40^{\mathrm{m}}35\overset{\mathrm{s}}{.}30$	$+56°24'20\overset{''}{.}22$	Djorgovski	Dec. 2004	10800 s
			Becker	Mar. 2006	1200 s
SDSS J0842+1218	$08^{\mathrm{h}}42^{\mathrm{m}}29\overset{\mathrm{s}}{.}43$	$+12°18'50\overset{''}{.}58$	Becker	Mar. 2006	2400 s
SDSS J0927+2001	$09^{\mathrm{h}}27^{\mathrm{m}}21\overset{\mathrm{s}}{.}82$	$+20°01'23\overset{''}{.}64$	Becker	Mar. 2006	1200 s
SDSS J1030+0524	$10^{\mathrm{h}}30^{\mathrm{m}}27\overset{\mathrm{s}}{.}11$	$+05°24'55\overset{''}{.}06$	Becker	May 2001	1800 s
			Becker	Jan. 2002	12000 s
			Madau	Feb. 2002	7339 s
			Cowie	Feb. 2002	16200 s
			Djorgovski	Mar. 2002	16200 s
			Cowie	Jan. 2003	5400 s
SDSS J1048+4637	$10^{\mathrm{h}}48^{\mathrm{m}}45\overset{\mathrm{s}}{.}07$	$+46°37'18\overset{''}{.}55$	Becker	Dec. 2002	3600 s
			Cowie	Jan. 2003	18000 s
			Cowie	Mar. 2003	5400 s
			Sanchez	Jun. 2003	5400 s
			Djorgovski	Dec. 2004	6600 s
SDSS J1137+3549	$11^{\mathrm{h}}37^{\mathrm{m}}17\overset{\mathrm{s}}{.}73$	$+35°49'56\overset{''}{.}85$	White	Jan. 2016	7800 s
			Becker	Jan. 2005	2400 s
SDSS J1148+5251	$11^{\mathrm{h}}48^{\mathrm{m}}16\overset{\mathrm{s}}{.}65$	$+52°51'50\overset{''}{.}39$	Becker	May 2002	8100 s
			Becker	Dec. 2002	22800 s
			Djorgovski	Dec. 2002	11400 s
			Kulkarni	Jan. 2003	3600 s
			Cowie	Jan. 2003	16200 s
			Becker	Feb. 2003	16800 s
			Cowie	Jan. 2004	11700 s
SDSS J1250+3130	$12^{\mathrm{h}}50^{\mathrm{m}}51\overset{\mathrm{s}}{.}93$	$+31°30'21\overset{''}{.}90$	Becker	Jan. 2005	3600 s
SDSS J1306+0356	$13^{\mathrm{h}}06^{\mathrm{m}}08\overset{\mathrm{s}}{.}27$	$+03°56'26\overset{''}{.}36$	Becker	May 2001	900 s
			Cowie	Feb. 2002	15300 s
			Djorgovski	Mar. 2002	12600 s
			Cowie	Mar. 2003	9000 s
			Cowie	Mar. 2004	12600 s
ULAS J1319+0950	$13^{\mathrm{h}}19^{\mathrm{m}}11\overset{\mathrm{s}}{.}30$	$+09°50'51\overset{''}{.}52$	Steidel	Mar. 2008	2400 s
SDSS J1335+3533	$13^{\mathrm{h}}35^{\mathrm{m}}50\overset{\mathrm{s}}{.}81$	$+35°33'15\overset{''}{.}82$	Becker	Mar. 2006	1200 s
SDSS J1411+1217	$14^{\mathrm{h}}11^{\mathrm{m}}11\overset{\mathrm{s}}{.}29$	$+12°17'37\overset{''}{.}28$	Djorgovski	Apr. 2004	10200 s
			Becker	Jan. 2005	3600 s

			Cowie	May 2005	10800 s
			Cowie	Mar. 2008	12600 s
SDSS J1602+4228	$16^\mathrm{h}02^\mathrm{m}53\overset{s}{.}98$	$+42°28'24\overset{''}{.}94$	Cowie	May. 2005	10620 s
			Cowie	Aug. 2005	3600 s
SDSS J1623+3112	$16^\mathrm{h}23^\mathrm{m}31\overset{s}{.}81$	$+31°12'00\overset{''}{.}53$	Becker	Jan. 2004	3600 s
SDSS J1630+4012	$16^\mathrm{h}30^\mathrm{m}33\overset{s}{.}90$	$+40°12'09\overset{''}{.}69$	Sanchez	Jun. 2003	17700 s
			Sargent	Oct. 2010	2000 s
CFHQS J1641+3755	$16^\mathrm{h}41^\mathrm{m}21\overset{s}{.}73$	$+37°55'20\overset{''}{.}15$	Willott	Oct. 2007	2400 s
SDSS J2054-0005	$20^\mathrm{h}54^\mathrm{m}06\overset{s}{.}49$	$-00°05'14\overset{''}{.}80$	Sargent	Oct. 2010	12000 s
CFHQS J2229+1457	$22^\mathrm{h}29^\mathrm{m}01\overset{s}{.}65$	$+14°57'09\overset{''}{.}00$	Willott	Sep. 2008	3600 s
			Prochaska[*]	Sep. 2016	3600 s
SDSS J2315-0023	$23^\mathrm{h}15^\mathrm{m}46\overset{s}{.}57$	$-00°23'58\overset{''}{.}10$	Becker	Jan. 2006	1200 s
			Sargent	Oct. 2010	27000 s
CFHQS J2329-0301	$23^\mathrm{h}29^\mathrm{m}08\overset{s}{.}28$	$-03°01'58\overset{''}{.}80$	Willott	Oct. 2007	10800 s
			Willott	Sep. 2008	14400 s

Notes. Columns show the object name, its coordinates, the principal investigator of the different observing runs, the observation date, and total exposure time of the object in this run.

[*] spectrum taken with LRIS.

Table A.2 Catalog for the data release.

object	RA [deg]	DEC [deg]	z_{em}	M_{1450}	S/N	$R[\Delta\lambda/\lambda]$	telescope	instrument
SDSS J0002+2550	0.6641	25.843	5.82	-27.31	57	5400	Keck II	ESI
SDSS J0005-0006	1.4681	-0.1155	5.844	-25.73	13	4000	Keck II	ESI
CFHQS J0050+3445	13.7621	34.756	6.253	-26.7	18	4000	Keck II	ESI
SDSS J0100+2802	15.0542	28.0405	6.3258	-29.14	35	4000	Keck II	ESI
ULAS J0148+0600	27.1568	6.0056	5.98	-27.39	20	4000	Keck II	ESI
ULAS J0203+0012	30.8849	0.2081	5.72	-26.26	5	4000	Keck II	ESI
CFHQS J0210-0456	32.555	-4.9391	6.4323	-24.53	1	5400	Keck II	ESI
PSO J0226+0302	36.5078	3.0498	6.5412	-27.33	9	4000	Keck II	ESI
CFHQS J0227-0605	36.9304	-6.0917	6.2	-25.28	3	4000	Keck II	ESI
SDSS J0303-0019	45.8808	-0.3202	6.078	-25.56	2	4000	Keck II	ESI
SDSS J0353+0104	58.4572	1.068	6.072	-26.43	10	4000	Keck II	ESI
PSO J0402+2451	60.5529	24.8568	6.18	-26.95	11	4000	Keck II	ESI
SDSS J0818+1722	124.6142	17.3811	6.02	-27.52	5	4000	Keck II	ESI
SDSS J0836+0054	129.1827	0.9148	5.81	-27.75	108	4000	Keck II	ESI
SDSS J0840+5624	130.1471	56.4056	5.8441	-27.24	23	4000	Keck II	ESI
SDSS J0842+1218	130.6226	12.314	6.069	-26.91	8	4000	Keck II	ESI
SDSS J0927+2001	141.8409	20.0232	5.7722	-26.76	6	4000	Keck II	ESI
SDSS J1030+0524	157.613	5.4153	6.309	-26.99	20	4000	Keck II	ESI
SDSS J1048+4637	162.1878	46.6218	6.2284	-27.24	42	4000	Keck II	ESI
SDSS J1137+3549	174.3239	35.8325	6.03	-27.36	18	4000	Keck II	ESI
SDSS J1148+5251	177.0694	52.864	6.4189	-27.62	28	4000	Keck II	ESI
SDSS J1250+3130	192.7164	31.5061	6.15	-26.53	7	4000	Keck II	ESI

SDSS J1306+0356	196.5345	3.9907	6.016	-26.81	28	4000	Keck II	ESI
ULAS J1319+0950	199.7971	9.8476	6.133	-27.05	8	4000	Keck II	ESI
SDSS J1335+3533	203.9617	35.5544	5.9012	-26.67	5	4000	Keck II	ESI
SDSS J1411+1217	212.797	12.2937	5.904	-26.69	25	4000	Keck II	ESI
SDSS J1602+4228	240.7249	42.4736	6.09	-26.94	16	4000	Keck II	ESI
SDSS J1623+3112	245.8825	31.2001	6.2572	-26.55	9	4000	Keck II	ESI
SDSS J1630+4012	247.6412	40.2027	6.065	-26.19	11	4000	Keck II	ESI
CFHQS J1641+3755	250.3405	37.9223	6.047	-25.67	2	4000	Keck II	ESI
SDSS J2054-0005	313.527	-0.0874	6.0391	-26.21	12	4000	Keck II	ESI
CFHQS J2229+1457	337.2569	14.9525	6.1517	-24.78	2	4000	Keck II	ESI
SDSS J2315-0023	348.944	-0.3995	6.117	-25.66	12	4000	Keck II	ESI
CFHQS J2329-0301	352.2845	-3.033	6.417	-25.25	2	4000	Keck II	ESI

Notes. The columns show the object name, the coordinates RA and DEC of the quasar given in degrees, the emission redshift and the quasar's magnitude M_{1450}, the S/N of the data, the telescope and instrument with which the spectra are observed and the spectral resolution of the data. Note that in some cases we co-added data with different spectral resolution.
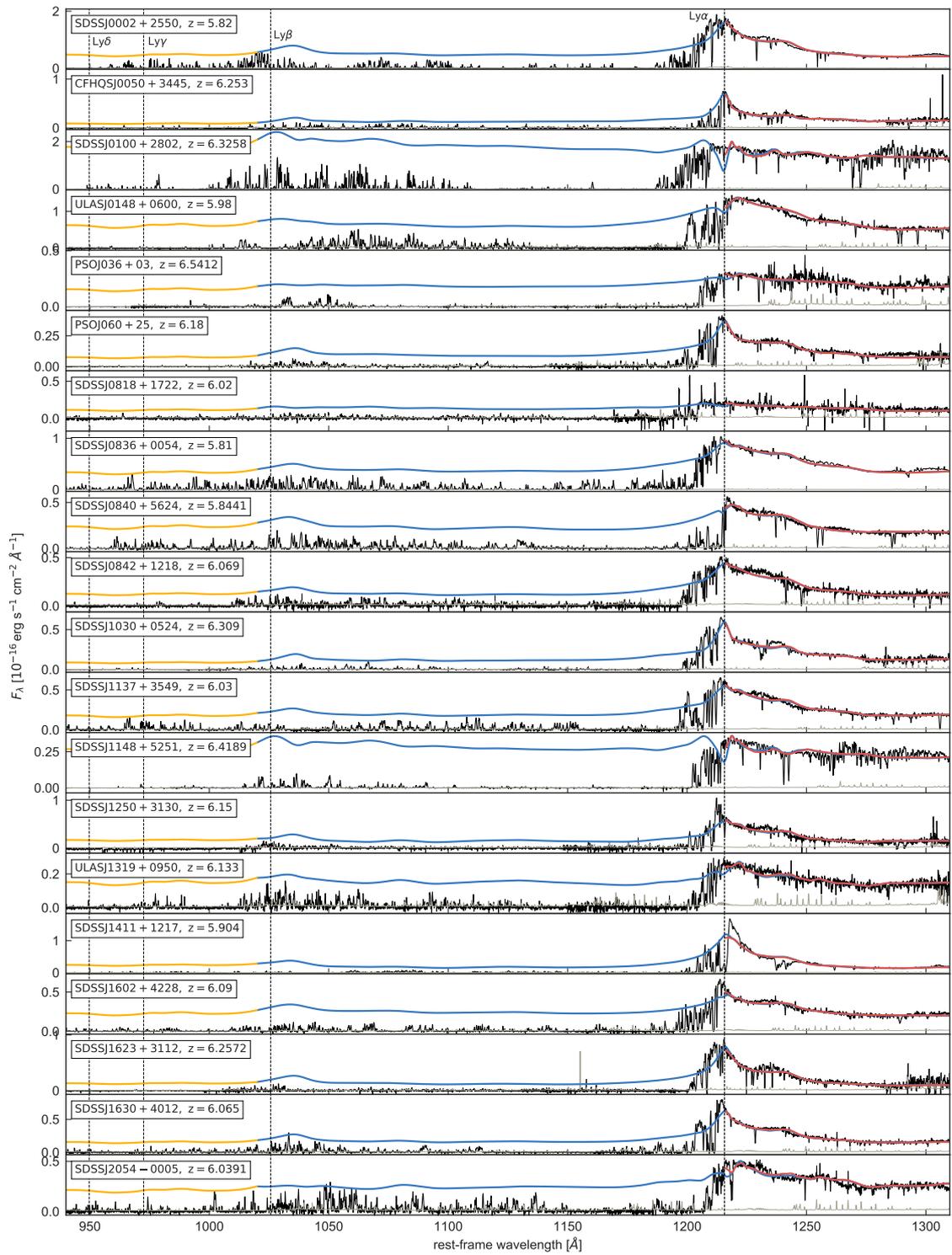
Figure A.1 Same as Fig. 8.3 for the remaining 20 quasar spectra in our data set.

The Correction of Zero-Level Offsets in Quasar Spectra

In order to correct for possible offsets in the zero-level of the quasar spectra, we examine the negative pixels in each spectral bin, which should appear to be a truncated Gaussian distribution, providing an estimate of the noise level in the spectra. To this end, we take all flux pixels F below zero in each spectral bin of $50\ \mathrm{cMpc}\,h^{-1}$, and calculate the CDF, ignoring correlations between neighboring pixels.

In the case of no offset in the zero-level within a spectral bin, the estimated mean of the CDF μ_{CDF} should be equal to zero, as shown by the toy model example in the lower middle panel of Fig. B.1. The upper panels show the PDF of the same respective toy model case. However, if the zero-level is slightly under- or overestimated (left and right panels in Fig. B.1, respectively), the estimated μ_{CDF} tracking the true zero-level will likewise be below or above zero. It is clear that we can obtain a handle on these systematic offsets by examining the purely negative pixels and fitting a truncated CDF model.

We use a MCMC approach making use of the implementation of the affine-invariant ensemble sampler emcee (Foreman-Mackey et al., 2013) to estimate the mean μ_{CDF} of the best fitting model CDF

$$\mathrm{CDF}_{\mathrm{model}}(F) = \frac{A_{\mathrm{CDF}}}{2}\left[1 + \mathrm{erf}\left(\frac{F - \mu_{\mathrm{CDF}}}{\sqrt{2}\,\sigma_{\mathrm{CDF}}}\right)\right], \tag{B.1}$$

while marginalizing over the width of the distribution σ_{CDF} and the scaling factor A_{CDF}. The likelihood function just maximizes the least squares between the CDF model and the measured CDF, i.e.

$$\ln \mathscr{L} = -0.5\,(\mathrm{CDF}_{\mathrm{model}} - \mathrm{CDF}_{\mathrm{data}})^2. \tag{B.2}$$
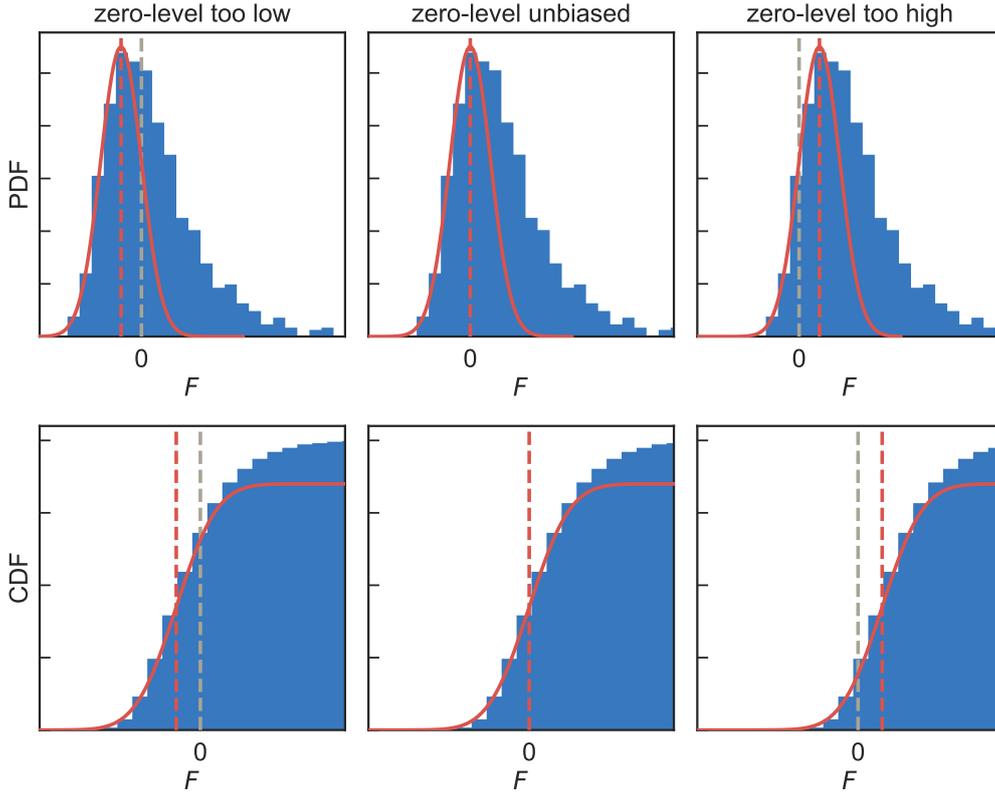
Figure B.1 Schematic depiction of the correction for offsets in the zero-level. The blue histograms show the PDF (top panels) and CDF (lower panels) of flux pixels from a toy model in a spectral bin. The red curves show a normal distribution of the expected noise level, with a mean indicated by the red dashed line. The grey dashed lines in the right and left panels show the current over- or underestimated zero-level, respectively.

We then take the median of the resulting posterior PDF as the best estimate for μ_{CDF}.

The three free parameters of the $\mathrm{CDF_{model}}$, μ_{CDF}, σ_{CDF}, and A_{CDF}, are highly degenerate with each other because we are fitting only a small part of the CDF when taking solely flux pixels below zero into account, i.e. $F < 0$. Thus we have to apply strict priors to break this degeneracy. The priors we chose are flat priors within the intervals $\mu_{\mathrm{CDF}} \in [-0.05, 0.05]$, since we expect the total offset of the zero-level to be less than $\pm 5\%$, $\sigma_{\mathrm{CDF}} \in [0.75\,\sigma_{\mathrm{eff}}, 1.25\,\sigma_{\mathrm{eff}}]$, which takes into account the noise vector at each pixel i of each quasar spectrum to estimate $\sigma_{\mathrm{eff}} = \sqrt{\sum_i \sigma_i^2/N}$, and $A_{\mathrm{CDF}} \in [0.45\,N_{F<0}, 0.55\,N_{F<0}]$. The upper and lower boundaries for A_{CDF} result from the fact that the number of pixels with flux below zero, i.e. $N_{F<0}$, in an unbiased case should be exactly half of the pixels, i.e. $0.5\,N_{F<0}$. In the presence of the possible offsets in the zero-level of the spectra we allow A_{CDF} to deviate from the unbiased case.
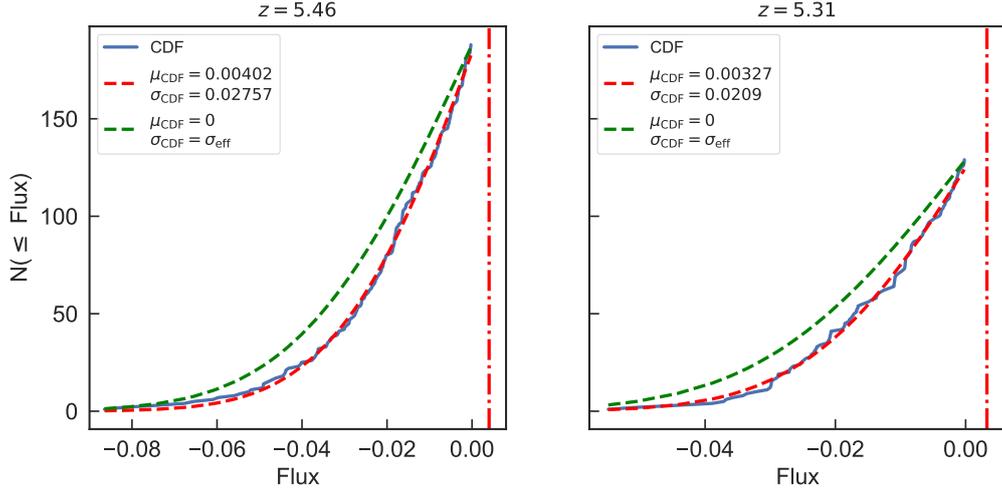
Figure B.2 Example CDF of two spectral bins along the sightline of SDSS J0840+5624. The blue curves show the measured CDF of each flux pixel, the red dotted curves show the best fitted CDF with the estimated values for mean and variance in the legend. The red dashed-dotted vertical line indicates the best fitted mean μ_{CDF}. The green dashed curves shows the ideal CDF, assuming no zero-level offset, i.e. $\mu_{\mathrm{CDF}} = 0$, and a correct noise model, i.e. $\sigma_{\mathrm{CDF}} = \sigma_{\mathrm{eff}}$.

Fig. B.2 shows two examples of the procedure. Both panels show the CDF for two spectral bins along the sightline of SDDS J0840+5624. We show the actually measured CDF of all negative flux pixels (blue curves) and overplot the best fitted CDF (red dashed curves, with mean μ_{CDF} indicated by the red dashed-dotted lines) and the ideal CDF with no zero-level offset, i.e. $\mu_{\mathrm{CDF}} = 0$, and the variance given by the noise of the data, i.e. $\sigma_{\mathrm{CDF}} = \sigma_{\mathrm{eff}}$. We can see that we obtain small zero-level offsets of about $\Delta\langle F^{\mathrm{Ly}\alpha}\rangle \sim \mu_{\mathrm{CDF}} \approx 0.3 - 0.4\%$.

In the end, we offset all pixels i within each spectral bin by the respective best estimate for μ_{CDF}, i.e. $F_{i,\mathrm{new}} = F_i - \mu_{\mathrm{CDF}}$, and calculate the mean flux and the opacity from the offseted pixels $F_{i,\mathrm{new}}$.
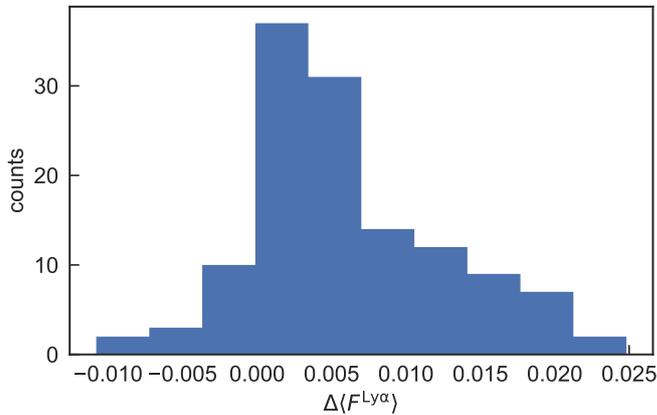
Figure B.3 Distribution of shifts in the mean flux estimate $\Delta\langle F^{\mathrm{Ly}\alpha}\rangle$ due to offsets in the zero-level of the quasar spectra.

Fig. B.3 shows the difference in mean flux estimates due to corrections in the zero-level. We estimate this systematic uncertainty in the mean flux measurements from calculating the 84th percentile and 16th percentile of this distribution and taking their average, which results in

$$\sigma_{\langle F^{\mathrm{Ly}\alpha}\rangle} = 0.0067.$$

Individual Opacity Measurements within the Lyα Forest

Our measurements of the mean flux in the Lyα forest are shown in Table C.1 for spectral bins of size 50 cMpc h^{-1} along all 23 quasar sight lines in our data sample. In Fig. C.1, C.2, and C.3 we show these spectral chunks, and indicate the respective optical depth measurements.

Table C.1   Mean flux measurements in the Lyα forest.

object	z_{start}	z_{abs}	z_{end}	$\langle F \rangle$	$\sigma_{\langle F \rangle}$
SDSS J0002+2550	5.54	5.62	5.70	0.0472	0.0004
	5.39	5.47	5.54	0.0315	0.0005
	5.25	5.32	5.39	0.0241	0.0004
	5.10	5.17	5.25	0.1170	0.0005
	4.97	5.04	5.10	0.1239	0.0006
SDSS J0005-0006	5.60	5.68	5.76	0.0507	0.0016
	5.45	5.53	5.60	0.0384	0.0028
	5.30	5.38	5.45	0.0484	0.0028
	5.16	5.23	5.30	0.0460	0.0023
	5.02	5.09	5.16	0.1128	0.0031
	4.89	4.95	5.02	0.0732	0.0028
CFHQS J0050+3445	5.96	6.05	6.14	0.0045	0.0018
	5.80	5.88	5.96	0.0217	0.0031
	5.64	5.72	5.80	0.0269	0.0022
	5.49	5.56	5.64	0.0570	0.0020
	5.34	5.41	5.49	0.1336	0.0020
	5.19	5.26	5.34	0.1009	0.0014
SDSS J0100+2802	5.81	5.89	5.98	0.0013	0.0005
	5.65	5.73	5.81	0.0038	0.0003

	5.50	5.57	5.65	0.0416	0.0003
	5.35	5.42	5.50	0.0778	0.0003
ULAS J0148+0600[a]	5.68	5.76	5.84	0.0056	0.0018
	5.53	5.60	5.68	-0.0020	0.0016
	5.38	5.45	5.53	0.0533	0.0018
	5.23	5.30	5.38	0.0826	0.0014
	5.09	5.16	5.23	0.1428	0.0014
	4.95	5.02	5.09	0.1340	0.0018
PSO J036+03	6.24	6.33	6.42	-0.0018	0.0015
	6.06	6.15	6.24	-0.0032	0.0011
	5.89	5.98	6.06	-0.0021	0.0009
	5.73	5.81	5.89	0.0001	0.0008
	5.57	5.65	5.73	0.0065	0.0006
	5.42	5.50	5.57	0.0632	0.0012
PSO J060+25	5.90	5.98	6.06	0.0308	0.0037
	5.73	5.81	5.90	0.0154	0.0054
	5.57	5.65	5.73	0.0176	0.0020
	5.42	5.50	5.57	0.0256	0.0023
	5.28	5.35	5.42	0.0155	0.0017
	5.13	5.20	5.28	0.0646	0.0016
SDSS J0818+1722	5.56	5.64	5.72	0.0118	0.0034
	5.41	5.49	5.56	0.0194	0.0042
	5.26	5.34	5.41	0.0462	0.0032
	5.12	5.19	5.26	0.0457	0.0028
	4.99	5.05	5.12	0.0768	0.0041
SDSS J0836+0054	5.54	5.62	5.70	0.1061	0.0003
	5.39	5.46	5.54	0.0997	0.0004
	5.24	5.31	5.39	0.0303	0.0003
	5.10	5.17	5.24	0.1104	0.0003
	4.97	5.03	5.10	0.1036	0.0004
	4.83	4.90	4.97	0.1673	0.0003
SDSS J0840+5624	5.39	5.46	5.54	0.0279	0.0012
	5.24	5.31	5.39	0.0778	0.0009
	5.10	5.17	5.24	0.1010	0.0011
	4.96	5.03	5.10	0.0937	0.0014
	4.83	4.90	4.96	0.1487	0.0008
SDSS J0842+1218	5.76	5.84	5.92	-0.0177	0.0055
	5.60	5.68	5.76	0.0160	0.0037
	5.44	5.52	5.60	0.0285	0.0052
	5.30	5.37	5.44	0.0817	0.0048
	5.15	5.22	5.30	0.1302	0.0041
	5.02	5.08	5.15	0.0311	0.0046
SDSS J1030+0524	5.98	6.07	6.16	0.0055	0.0013
	5.82	5.90	5.98	0.0140	0.0020

	5.66	5.74	5.82	0.0114	0.0011
	5.50	5.58	5.66	0.0444	0.0010
	5.35	5.43	5.50	0.1219	0.0012
	5.21	5.28	5.35	0.0561	0.0008
SDSS J1137+3549	5.71	5.79	5.87	0.0056	0.0025
	5.56	5.63	5.71	0.0807	0.0017
	5.40	5.48	5.56	0.1422	0.0020
	5.26	5.33	5.40	0.0913	0.0017
	5.12	5.19	5.26	0.1193	0.0017
SDSS J1148+5251	5.77	5.86	5.94	0.0022	0.0003
	5.61	5.69	5.77	0.0109	0.0001
	5.46	5.54	5.61	0.0123	0.0002
	5.31	5.39	5.46	0.0422	0.0002
SDSS J1250+3130	5.83	5.91	5.99	-0.0212	0.0078
	5.67	5.75	5.83	-0.0091	0.0054
	5.51	5.59	5.67	-0.0099	0.0034
	5.36	5.44	5.51	0.0037	0.0042
	5.22	5.29	5.36	0.0276	0.0030
	5.08	5.15	5.22	0.0644	0.0027
SDSS J1306+0356	5.72	5.80	5.89	0.0934	0.0013
	5.57	5.64	5.72	0.0495	0.0007
	5.42	5.49	5.57	0.0491	0.0008
	5.27	5.34	5.42	0.0436	0.0007
	5.13	5.20	5.27	0.0630	0.0006
	4.99	5.06	5.13	0.0785	0.0008
ULAS J1319+0950	5.86	5.94	6.02	0.0019	0.0058
	5.70	5.78	5.86	0.0002	0.0049
	5.54	5.62	5.70	0.0249	0.0029
	5.39	5.46	5.54	0.0697	0.0036
	5.24	5.31	5.39	0.0406	0.0027
	5.10	5.17	5.24	0.1396	0.0029
SDSS J1411+1217	5.63	5.71	5.79	0.0170	0.0008
	5.48	5.56	5.63	0.0355	0.0010
	5.33	5.40	5.48	0.0543	0.0010
	5.19	5.26	5.33	0.0384	0.0008
	5.05	5.12	5.19	0.1211	0.0010
	4.91	4.98	5.05	0.0429	0.0009
SDSS J1602+4228	5.76	5.85	5.93	0.0242	0.0028
	5.61	5.68	5.76	0.0273	0.0013
	5.45	5.53	5.61	0.0381	0.0014
	5.30	5.38	5.45	0.0504	0.0012
	5.16	5.23	5.30	0.0609	0.0009
	5.02	5.09	5.16	0.0660	0.0011
SDSS J1623+3112	5.95	6.04	6.12	0.0136	0.0031

	z_{start}	z_{abs}	z_{end}	flux	unc.
	5.63	5.71	5.79	0.0276	0.0030
	5.47	5.55	5.63	0.0383	0.0034
	5.32	5.40	5.47	0.0198	0.0029
	5.18	5.25	5.32	0.0225	0.0018
SDSS J1630+4012	5.62	5.70	5.78	-0.0062	0.0023
	5.47	5.54	5.62	0.0193	0.0024
	5.32	5.39	5.47	0.1088	0.0028
	5.17	5.24	5.32	0.0586	0.0020
	5.04	5.10	5.17	0.1274	0.0028
SDSS J2054-0005	5.78	5.86	5.95	0.0287	0.0039
	5.62	5.70	5.78	0.0150	0.0026
	5.47	5.54	5.62	0.1131	0.0031
	5.32	5.39	5.47	0.0652	0.0032
	5.18	5.25	5.32	0.1196	0.0025
	5.04	5.11	5.18	0.2436	0.0037
SDSS J2315-0023	5.84	5.93	6.01	-0.0127	0.0041
	5.53	5.60	5.68	0.0477	0.0037
	5.38	5.45	5.53	0.0348	0.0042
	5.23	5.30	5.38	0.1012	0.0031
	5.09	5.16	5.23	0.1468	0.0031

Notes. The different columns show the name of the object, the beginning of each redshift bin z_{start}, the mean redshift of each bin z_{abs} and the end of z_{end} the redshift bin, and the mean flux of the continuum normalized spectrum with its uncertainty.

[a] Note that the measurements along this quasar sightline have been replaced by the ones from Becker et al. (2015b) in our master compilation.
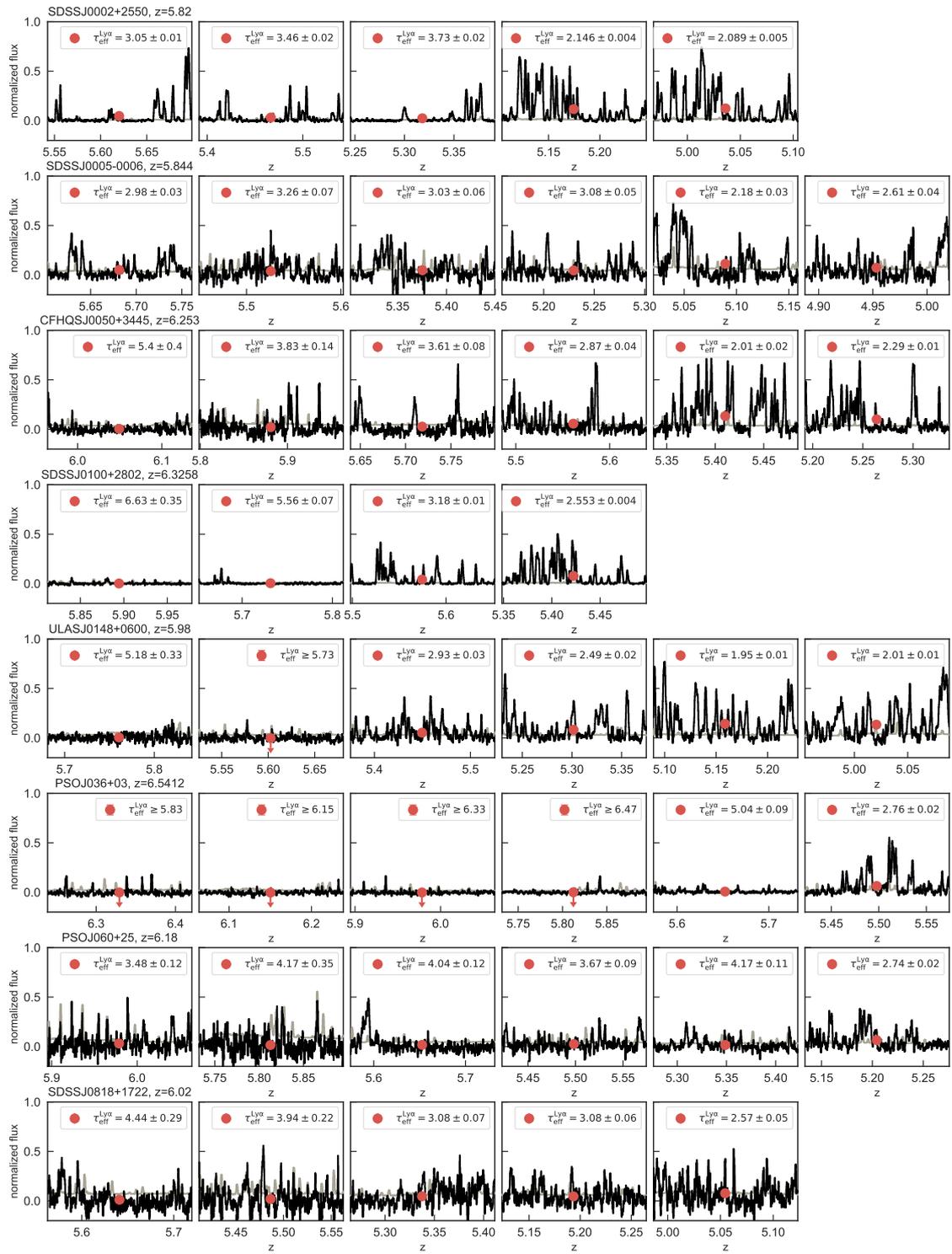
Figure C.1 Spectral bins of $50\,\mathrm{cMpc}\,h^{-1}$ along all quasar sightlines in our data sample, for which we measure $\tau_{\mathrm{eff}}^{\mathrm{Ly}\alpha}$ within the Lyα forest. The red data points show the measurements of $\langle F^{\mathrm{obs}} \rangle$ and the corresponding optical depth measurements are shown in the legend.
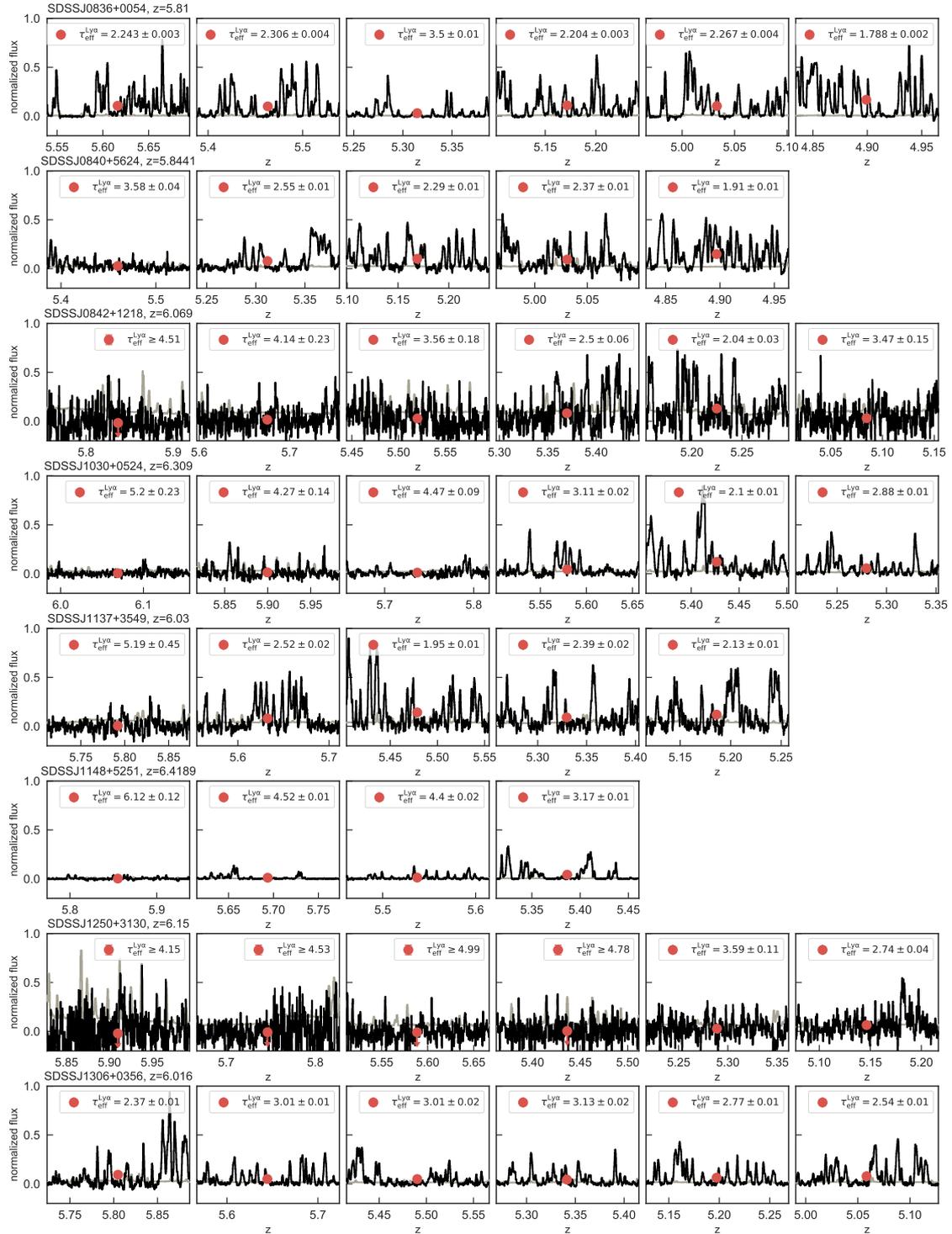
**Figure C.2**   Continuation of Fig. C.1.
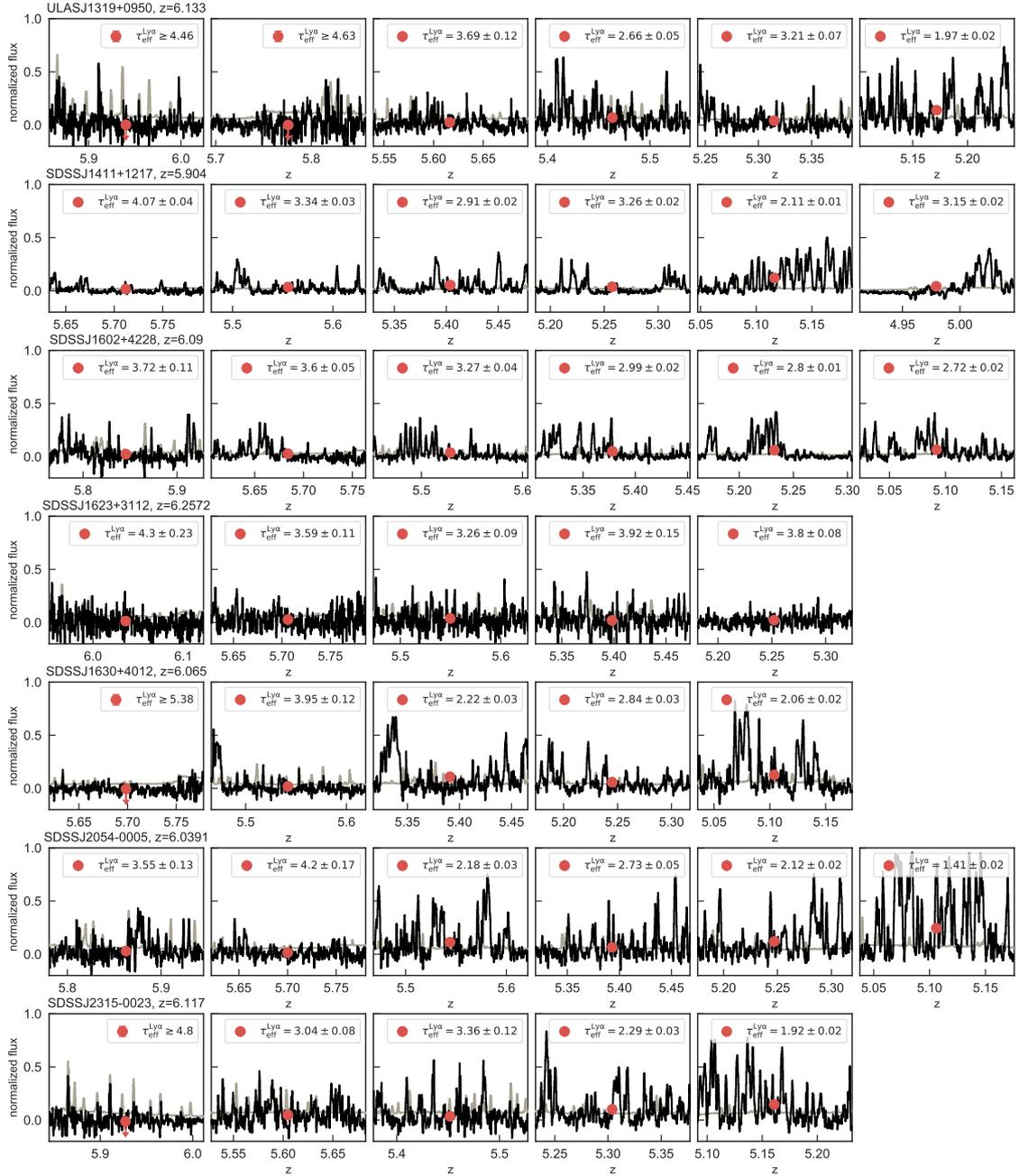
Figure C.3 Continuation of Fig. C.1 and Fig. C.2.

Associated Absorption Systems in SDSS J0840+5624

The quasar SDSS J0840 + 5624 has been excluded from our analysis because we found several absorption systems at high redshift associated with the quasar itself. These absorbers cause additional absorption within the proximity zone of the quasar, such that our analysis recovers a spuriously small proximity zone of $R_p \approx 0.88$ pMpc. However, inspecting the spectrum by eye reveals that the actual proximity zone of this object extends much further out to $R_p \gtrsim 5.0$ pMpc.

We find two N V absorption line systems ($\lambda_{\mathrm{rest}} = 1238.82$ Å and $\lambda_{\mathrm{rest}} = 1242.80$ Å) associated with close absorbers at $z \approx 5.835$ and $z \approx 5.825$, which are shown in Fig. D.1. We also find another absorption system in the spectrum at $z \approx 5.594$ with associated Si II absorption ($\lambda_{\mathrm{rest}} = 1260.42$ Å) that falls within the proximity zone. Thus there is a lot of absorption within the proximity zone that cannot be contributed to residual neutral hydrogen only and an exact determination of the proximity zone for this object without contamination from other absorption systems is impossible.
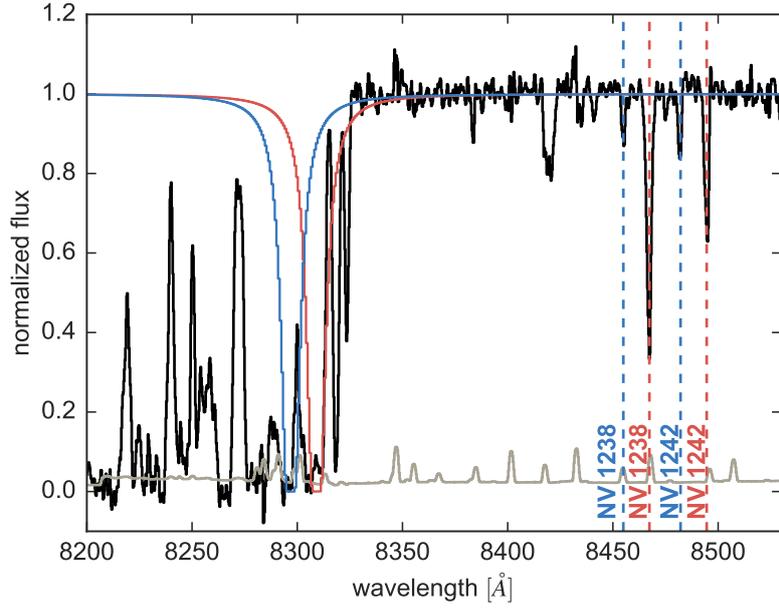
Figure D.1 Two N V absorption line systems shown as the blue and red dashed lines associated with close absorbers (blue and red curves) at $z_{\mathrm{blue}} \approx 5.825$ and $z_{\mathrm{red}} \approx 5.835$ in the spectrum of the quasar SDSS J0840 + 5624. The continuum normalized spectrum (black curve) and its noise vector (gray curve) are smoothed with a three pixel boxcar function (Note that in this figure we show the quasar spectrum normalized with a hand-fitted continuum.). The depicted absorbers have a column density $N_{\mathrm{HI}} = 10^{19}\mathrm{cm}^{-2}$ and a Doppler parameter $b = 20$ km/s.

Glossary

AGN active galactic nucleus

ALMA Atacama Large Millimetre Array

BAL broad absorption line

BLR broad line region

CDF cumulative distribution function

CFHQS Canada-France High-z Quasar Survey

CGM circumgalactic medium

CMB Cosmic Microwave Background

COS Cosmic Origins Spectrograph

DEIMOS Deep Imaging Multi-Object Spectrograph

DLA Damped Lyα System

ENLR Extended Narrow Line Region

ESI Echellette Spectrograph and Imager

FWHM Full Width Half Maximum

GNIRS Gemini Near Infra-Red Spectrograph

GP Gunn-Peterson

HST Hubble Space Telescope

HUDF Hubble Ultra Deep Field

IAU International Astronomical Union

IDL Interactive Data Language

IFU Integral Field Unit

IGM intergalactic medium

IR infra-red

JWST James Webb Space Telescope

KOA Keck Observatory Archive

LLAMAS Large Lenslet Array Magellan Spectrograph

LLS Lyman Limit system

LRIS Low Resolution Imaging Spectrometer

MCMC Markov Chain Monte Carlo

MUSE Multi Unit Spectroscopic Explorer

NFW Navarro-Frenk-White

NIR near infra-red

NOEMA Northern Extended Millimeter Array

PCA Principal Component Analysis

PCS principal component spectrum

PDF probability distribution function

PI principal investigator

PS1 Pan-STARRS 1

RGB red giant branch

SDSS Sloan Digital Sky Survey

SED spectral energy distribution

SMBH supermassive black hole

UV ultraviolet

UVB ultraviolet background

VLBA Very Long Baseline Array

VLT Very Large Telescope

WLQ weak emission line quasar

Publications of A.-C. Eilers

Parts of the work presented in this thesis have previously appeared in the following articles:

- **A.-C. Eilers**, F. B. Davies, J. F. Hennawi, J. X. Prochaska, Z. Lukić, & C. Mazzucchelli (2017a), *Implications of z ~ 6 Quasar Proximity Zones for the Epoch of Reionization and Quasar Lifetimes*, Astrophysical Journal, 840, 24–48

- **A.-C. Eilers**, F. B. Davies, & J. F. Hennawi (2018a), *The Opacity of the Intergalactic Medium Measured Along Quasar Sightlines at z ~ 6*, Astrophysical Journal, 864, 53–75

- **A.-C. Eilers**, J. F. Hennawi, & F. B. Davies (2018b), *First Spectroscopic Study of a Young Quasar*, Astrophysical Journal, 867, 30–39

- **A.-C. Eilers**, D. W. Hogg, H.-W. Rix, & M. Ness (2019a), *The Circular Velocity Curve of the Milky Way from 5 to 25 kpc*, Astrophysical Journal, 871, 120–128

- D. W. Hogg, **A.-C. Eilers**, & H.-W. Rix (2019), *Spectrophotometric parallaxes with linear models: Accurate distances for luminous red-giant stars*, submitted to the Astrophysical Journal (arXiv: 1810.09468)

- **A.-C. Eilers**, J. F. Hennawi, F. B. Davies, & J. Oñorbe (2019b), *Anomaly in the Opacity of the Intergalactic Medium in the Lyα and Lyβ Forest*, to be submitted to the Astrophysical Journal

Other publications including A.-C. Eilers (as first author or co-author):

- F. B. Davies, J. F. Hennawi, & **A.-C. Eilers** (2019), *Radiatively inefficient growth of supermassive black holes during the epoch of reionization*, submitted to Science

- K. Kakiichi, R. S. Ellis, N. Laporte, A. Zitrin, **A.-C. Eilers** et al. (2018), *The Role of Galaxies and AGN in Reionising the IGM - I: Keck Spectroscopy*

of 5 < z < 7 Galaxies in the QSO Field J1148+5251, Monthly Notices of the Royal Astronomical Society, 479, 43–63

- F. B. Davies, J. F. Hennawi, **A.-C. Eilers**, & Z. Lukić (2018), *A New Method to Measure the Post-Reionization Ionizing Background from the Joint Distribution of Lyman-α and Lyman-β Forest Transmission*, Astrophysical Journal, 855, 106–120

- K. El-Badry, H.-W. Rix, Y.-S. Ting, D. R. Weisz, M. Bergemann, P. Cargile, C. Conroy, & **A.-C. Eilers** (2018), *Signatures of unresolved binaries in stellar spectra: implications for spectral fitting*, Monthly Notices of the Royal Astronomical Society, 473, 5043–5049

- **A.-C. Eilers**, J. F. Hennawi, & K.-G. Lee (2017b), *Joint Bayesian Estimation of Quasar Continua and the Lyman-Alpha Forest Flux Probability Distribution Function*, Astrophysical Journal, 844, 136–160

- C. Mazzucchelli, E. Bañados, B. P. Venemans, R. Decarli, E. P. Farina, F. Walter, **A.-C. Eilers** et al. (2017) *Physical properties of 15 quasars at $z \gtrsim 6.5$*, Astrophysical Journal, 849, 91–117

- K.-G. Lee, J. F. Hennawi, C. Stark, J. X. Prochaska, M. White, D. J. Schlegel, **A.-C. Eilers** et al. (2014), *Lyman-alpha Forest Tomography from Background Galaxies: The First Megaparsec-Resolution Large-Scale Structure Map at z>2*, Astrophysical Journal Letters, 795, 12–19

- J. H. J. de Bruijne & **A.-C. Eilers** (2012), *Radial velocities for the Hipparcos-Gaia Hundred-Thousand-Proper-Motion project*, Astronomy & Astrophysics, 546, 61–75

Bibliography

Ablimit, I., & Zhao, G. 2017, ApJ, 846, 10

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

Alpher, R. A., Bethe, H., & Gamow, G. 1948, Physical Review, 73, 803

Antoja, T., Roca-Fàbrega, S., de Bruijne, J., & Prusti, T. 2016, A&A, 589, A13

Antoja, T., Helmi, A., Romero-Gómez, M., et al. 2018, Nature, 561, 360

Antonucci, R. 1993, ARA&A, 31, 473

Bañados, E., Carilli, C., Walter, F., et al. 2018a, ApJL, 861, L14

Bañados, E., Decarli, R., Walter, F., et al. 2015, ApJL, 805, L8

Bañados, E., Venemans, B. P., Decarli, R., et al. 2016, ApJS, 227, 11

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

Bailer-Jones, C. A. L., Rybizki, J., Fouesneau, M., Mantelet, G., & Andrae, R. 2018, AJ, 156, 58

Bajtlik, S., Duncan, R. C., & Ostriker, J. P. 1988, ApJ, 327, 570

Bardeen, J. M., Steinhardt, P. J., & Turner, M. S. 1983, PhRvD, 28, 679

Barnett, R., Warren, S. J., Becker, G. D., et al. 2017, A&A, 601, A16

Becker, G. D., Bolton, J. S., & Lidz, A. 2015a, PASA, 32, e045

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

Becker, G. D., Davies, F. B., Furlanetto, S. R., et al. 2018, ApJ, 863, 92

Becker, G. D., Sargent, W. L. W., Rauch, M., & Calverley, A. P. 2011, ApJ, 735, 93

Becklin, E. E., & Neugebauer, G. 1968, ApJ, 151, 145

Belokurov, V. 2013, NewAR, 57, 100

Bensby, T., Feltzing, S., & Oey, M. S. 2014, A&A, 562, A71

Bernstein, R. M., Burles, S. M., & Prochaska, J. X. 2015, PASP, 127, 911

Bhattacharjee, P., Chaudhury, S., & Kundu, S. 2014, ApJ, 785, 63

Bienaymé, O., Famaey, B., Siebert, A., et al. 2014, A&A, 571, A92

Binney, J., Gerhard, O., & Spergel, D. 1997, MNRAS, 288, 365

Binney, J., & Tremaine, S. 2008, Galactic Dynamics: Second Edition (Princeton University Press)

Bland-Hawthorn, J., & Gerhard, O. 2016, ARA&A, 54, 529

Bochanski, J. J., Hennawi, J. F., Simcoe, R. A., et al. 2009, PASP, 121, 1409

Bolton, J. S., Becker, G. D., Raskutti, S., et al. 2012, MNRAS, 419, 2880

Bolton, J. S., Becker, G. D., Wyithe, J. S. B., Haehnelt, M. G., & Sargent, W. L. W. 2010, MNRAS, 406, 612

Bolton, J. S., & Haehnelt, M. G. 2007a, MNRAS, 381, L35

Bolton, J. S., & Haehnelt, M. G. 2007b, MNRAS, 374, 493

Bolton, J. S., & Haehnelt, M. G. 2007c, MNRAS, 382, 325

Bolton, J. S., Haehnelt, M. G., Warren, S. J., et al. 2011, MNRAS, 416, L70

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

Bovy, J. 2017, MNRAS, 468, L63

Bovy, J., & Rix, H.-W. 2013, ApJ, 779, 115

Bovy, J., Allende Prieto, C., Beers, T. C., et al. 2012, ApJ, 759, 131

Boylan-Kolchin, M., Bullock, J. S., & Kaplinghat, M. 2011, MNRAS, 415, L40

Boylan-Kolchin, M., Bullock, J. S., & Kaplinghat, M. 2012, MNRAS, 422, 1203

Brand, J., & Blitz, L. 1993, A&A, 275, 67

Bullock, J. S., & Boylan-Kolchin, M. 2017, ARA&A, 55, 343

Cai, Z., Fan, X., Yang, Y., et al. 2017, ApJ, 837, 71

Cai, Z., Hamden, E., Matuszewski, M., et al. 2018, ApJL, 861, L3

Calverley, A. P., Becker, G. D., Haehnelt, M. G., & Bolton, J. S. 2011, MNRAS, 412, 2543

Cantalupo, S., Arrigoni-Battaia, F., Prochaska, J. X., Hennawi, J. F., & Madau, P. 2014, Nature, 506, 63

Carilli, C. L., Wang, R., Fan, X., et al. 2010, ApJ, 714, 834

Cen, R., & Haiman, Z. 2000, ApJ, 542, L75

Cen, R., Miralda-Escudé, J., Ostriker, J. P., & Rauch, M. 1994, ApJL, 437, L9

Cen, R., & Safarzadeh, M. 2015, ApJL, 798, L38

Chardin, J., Haehnelt, M. G., Aubert, D., & Puchwein, E. 2015, MNRAS, 453, 2943

Chardin, J., Puchwein, E., & Haehnelt, M. G. 2017, MNRAS, 465, 3429

Choi, E., Ostriker, J. P., Naab, T., & Johansson, P. H. 2012, ApJ, 754, 125

Churchwell, E., Babler, B. L., Meade, M. R., et al. 2009, PASP, 121, 213

Ciotti, L., Ostriker, J. P., & Proga, D. 2010, ApJ, 717, 708

Clowe, D., Bradač, M., Gonzalez, A. H., et al. 2006, ApJL, 648, L109

Colless, M., Dalton, G., Maddox, S., et al. 2001, MNRAS, 328, 1039

Conroy, C., & White, M. 2013, ApJ, 762, 70

Cooke, R. J., Pettini, M., & Steidel, C. C. 2017, MNRAS, 467, 802

Cooper, T. J., Simcoe, R. A., Cooksey, K. L., O'Meara, J. M., & Torrey, P. 2015, ApJ, 812, 58

Cracco, V., Ciroi, S., di Mille, F., et al. 2011, MNRAS, 418, 2630

Crighton, N. H. M., O'Meara, J. M., & Murphy, M. T. 2016, MNRAS, 457, L44

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

D'Aloisio, A., McQuinn, M., Davies, F. B., & Furlanetto, S. R. 2018, MNRAS, 473, 560

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

Davé, R., Hernquist, L., Katz, N., & Weinberg, D. H. 1999, ApJ, 511, 521

Davies, F. B., Becker, G. D., & Furlanetto, S. R. 2018a, ApJ, 860, 155

Davies, F. B., & Furlanetto, S. R. 2016, MNRAS, 460, 1328

Davies, F. B., Furlanetto, S. R., & McQuinn, M. 2016, MNRAS, 457, 3006

Davies, F. B., Hennawi, J. F., Eilers, A.-C., & Lukić, Z. 2018b, ApJ, 855, 106

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

De Rosa, G., Decarli, R., Walter, F., et al. 2011, ApJ, 739, 56

De Rosa, G., Venemans, B. P., Decarli, R., et al. 2014, ApJ, 790, 145

Dekel, A., Birnboim, Y., Engel, G., et al. 2009, Nature, 457, 451

Diamond-Stanic, A. M., Fan, X., Brandt, W. N., et al. 2009, ApJ, 699, 782

DiPompeo, M. A., Hickox, R. C., Myers, A. D., & Geach, J. E. 2017, MNRAS, 464, 3526

Dodelson, S. 2003, Modern cosmology

Draine, B. T. 2011, Physics of the Interstellar and Intergalactic Medium

Du, P., Hu, C., Lu, K.-X., et al. 2015, ApJ, 806, 22

Eadie, G., & Jurić, M. 2018, ArXiv e-prints, arXiv:1810.10036

Eadie, G. M., & Harris, W. E. 2016, ApJ, 829, 108

Eckart, A., Genzel, R., Hofmann, R., Sams, B. J., & Tacconi-Garman, L. E. 1995, ApJL, 445, L23

Eddington, A. S. 1922, MNRAS, 83, 32

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

Eilers, A.-C., Davies, F. B., Hennawi, J. F., et al. 2017a, ApJ, 840, 24

Eilers, A.-C., Hennawi, J. F., & Davies, F. B. 2018b, ApJ, 867, 30

Eilers, A.-C., Hennawi, J. F., & Lee, K.-G. 2017b, ApJ, 844, 136

Eilers, A.-C., Hogg, D. W., Rix, H.-W., & Ness, M. K. 2019, ApJ, 871, 120

Fall, S. M., & Efstathiou, G. 1980, MNRAS, 193, 189

Fan, X. 2006, New Astronomy Reviews, 50, 665

Fan, X., Carilli, C. L., & Keating, B. 2006a, ARA&A, 44, 415

Fan, X., White, R. L., Davis, M., et al. 2000, AJ, 120, 1167

Fan, X., Hennawi, J. F., Richards, G. T., et al. 2004, AJ, 128, 515

Fan, X., Strauss, M. A., Richards, G. T., et al. 2006b, AJ, 131, 1203

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

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

Fich, M., Blitz, L., & Stark, A. A. 1989, ApJ, 342, 272

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

Frenk, C. S., & White, S. D. M. 2012, Annalen der Physik, 524, 507

Fumagalli, M., O'Meara, J. M., & Prochaska, J. X. 2011, Science, 334, 1245

Fumagalli, M., O'Meara, J. M., Prochaska, J. X., & Worseck, G. 2013, ApJ, 775, 78

Furlanetto, S. R., & Oh, S. P. 2009, ApJ, 701, 94

Gaia Collaboration, Prusti, T., de Bruijne, J. H. J., et al. 2016, A&A, 595, A1

Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2018, A&A, 616, A1

García Pérez, A. E., Allende Prieto, C., Holtzman, J. A., et al. 2016, AJ, 151, 144

Gaskell, C. M. 1982, ApJ, 263, 79

Geller, M. J., & Huchra, J. P. 1989, Science, 246, 897

Genzel, R., Schreiber, N. M. F., Übler, H., et al. 2017, Nature, 543, 397

Ghez, A. M., Klein, B. L., Morris, M., & Becklin, E. E. 1998, ApJ, 509, 678

Gnedin, N. Y., Becker, G. D., & Fan, X. 2017, ApJ, 841, 26

Gott, III, J. R., Jurić, M., Schlegel, D., et al. 2005, ApJ, 624, 463

Gravity Collaboration, Abuter, R., Amorim, A., et al. 2018, A&A, 615, L15

Gunn, J. E., Knapp, G. R., & Tremaine, S. D. 1979, AJ, 84, 1181

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

Guth, A. H. 1981, PhRvD, 23, 347

Guth, A. H., & Pi, S.-Y. 1982, Physical Review Letters, 49, 1110

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

Haiman, Z., & Cen, R. 2001, in Astronomical Society of the Pacific Conference Series, Vol. 222, The Physics of Galaxy Formation, ed. M. Umemura & H. Susa, 101

Haiman, Z., & Hui, L. 2001, ApJ, 547, 27

Hawking, S. W. 1982, Physics Letters B, 115, 295

Haywood, M., Di Matteo, P., Lehnert, M. D., Katz, D., & Gómez, A. 2013, A&A, 560, A109

Helmi, A., Babusiaux, C., Koppelman, H. H., et al. 2018, Nature, 563, 85

Hennawi, J. F., & Prochaska, J. X. 2007, ApJ, 655, 735

Hennawi, J. F., Prochaska, J. X., Cantalupo, S., & Arrigoni-Battaia, F. 2015, Science, 348, 779

Hennawi, J. F., Prochaska, J. X., Burles, S., et al. 2006, ApJ, 651, 61

Hernquist, L. 1992, ApJ, 400, 460

Hickox, R. C., & Alexander, D. M. 2018, ARA&A, 56, 625

Hogg, D. W., Eilers, A.-C., & Rix, H.-W. 2018, arXiv e-prints, arXiv:1810.09468

Hopkins, P. F., Hernquist, L., Cox, T. J., et al. 2005, ApJ, 630, 705

Hopkins, P. F., Hernquist, L., Cox, T. J., & Kereš, D. 2008, ApJS, 175, 356

Hopkins, P. F., Somerville, R. S., Cox, T. J., et al. 2009, MNRAS, 397, 802

Hryniewicz, K., Czerny, B., Nikołajuk, M., & Kuraszkiewicz, J. 2010, MNRAS, 404, 2028

Huang, Y., Liu, X.-W., Yuan, H.-B., et al. 2016, MNRAS, 463, 2623

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

Ishibashi, W., & Fabian, A. C. 2016, MNRAS, 463, 1291

Jeans, J. H. 1902, Philosophical Transactions of the Royal Society of London Series A, 199, 1

Jeans, J. H. 1915, MNRAS, 76, 70

Jiang, L., Fan, X., Vestergaard, M., et al. 2007, AJ, 134, 1150

Jiang, L., Fan, X., Annis, J., et al. 2008, AJ, 135, 1057

Johnson, J. L., Whalen, D. J., Agarwal, B., Paardekooper, J.-P., & Khochfar, S. 2014, MNRAS, 445, 686

Jurić, M., Ivezić, Ž., Brooks, A., et al. 2008, ApJ, 673, 864

Kafle, P. R., Sharma, S., Lewis, G. F., & Bland-Hawthorn, J. 2012, ApJ, 761, 98

Keating, L. C., Haehnelt, M. G., Cantalupo, S., & Puchwein, E. 2015, MNRAS, 454, 681

Keating, L. C., Puchwein, E., & Haehnelt, M. G. 2018, MNRAS, 477, 5501

Keel, W. C., Lintott, C. J., Maksym, W. P., et al. 2017, ApJ, 835, 256

Kelson, D. D. 2003, PASP, 115, 688

Khrykin, I. S., Hennawi, J. F., McQuinn, M., & Worseck, G. 2016, ApJ, 824, 133

Khrykin, I. S., Hennawi, J. F., & Worseck, G. 2018, arXiv e-prints, arXiv:1810.03391

Klypin, A., Kravtsov, A. V., Valenzuela, O., & Prada, F. 1999, ApJ, 522, 82

Klypin, A., Zhao, H., & Somerville, R. S. 2002, ApJ, 573, 597

Kollmeier, J. A., Zasowski, G., Rix, H.-W., et al. 2017, arXiv e-prints, arXiv:1711.03234

Koposov, S. E., Rix, H.-W., & Hogg, D. W. 2010, ApJ, 712, 260

Kormendy, J., & Ho, L. C. 2013, ARA&A, 51, 511

Kormendy, J., & Richstone, D. 1995, ARA&A, 33, 581

Kramer, R. H., & Haiman, Z. 2009, MNRAS, 400, 1493

Küpper, A. H. W., Balbinot, E., Bonaca, A., et al. 2015, ApJ, 803, 80

Kurk, J. D., Walter, F., Fan, X., et al. 2007, ApJ, 669, 32

LaMassa, S. M., Cales, S., Moran, E. C., et al. 2015, ApJ, 800, 144

Lang, P., Förster Schreiber, N. M., Genzel, R., et al. 2017, ApJ, 840, 92

Levine, E. S., Heiles, C., & Blitz, L. 2008, ApJ, 679, 1288

Liddle, A. 2003, An Introduction to Modern Cosmology, Second Edition, 188

Lidz, A., Hui, L., Zaldarriaga, M., & Scoccimarro, R. 2002, ApJ, 579, 491

Lidz, A., & Malloy, M. 2014, ApJ, 788, 175

Lidz, A., McQuinn, M., Zaldarriaga, M., Hernquist, L., & Dutta, S. 2007, ApJ, 670, 39

Lidz, A., Oh, S. P., & Furlanetto, S. R. 2006, ApJL, 639, L47

Lindegren, L., Hernández, J., Bombrun, A., et al. 2018, A&A, 616, A2

Liu, G., Zakamska, N. L., & Greene, J. E. 2014, MNRAS, 442, 1303

Liu, Y., & Zhang, S. N. 2011, ApJL, 728, L44

Lodato, G., & Natarajan, P. 2006, MNRAS, 371, 1813

Loeb, A., & Furlanetto, S. R. 2013, The First Galaxies in the Universe

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

Lusso, E., Worseck, G., Hennawi, J. F., et al. 2015, MNRAS, 449, 4204

Lynds, R. 1971, ApJL, 164, L73

Mackereth, J. T., Bovy, J., Leung, H. W., et al. 2019, arXiv e-prints, arXiv:1901.04502

Madau, P., & Dickinson, M. 2014, ARA&A, 52, 415

Madau, P., & Rees, M. J. 2000, ApJ, 542, L69

Maiolino, R., Cox, P., Caselli, P., et al. 2005, A&A, 440, L51

Majewski, S. R., Schiavon, R. P., Frinchaboy, P. M., et al. 2017, AJ, 154, 94

Mao, J., & Kim, M. 2016, ApJ, 828, 96

Margon, B., & Ostriker, J. P. 1973, ApJ, 186, 91

Marrese, P. M., Marinoni, S., Fabrizio, M., & Altavilla, G. 2018, arXiv e-prints, arXiv:1808.09151

Martini, P. 2004, Coevolution of Black Holes and Galaxies, 169

Martini, P., & Weinberg, D. H. 2001, ApJ, 547, 12

Maselli, A., Ferrara, A., & Gallerani, S. 2009, MNRAS, 395, 1925

Maselli, A., Gallerani, S., Ferrara, A., & Choudhury, T. R. 2007, MNRAS, 376, L34

Mazzucchelli, C., Bañados, E., Venemans, B. P., et al. 2017, ApJ, 849, 91

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

McElroy, R. E., Husemann, B., Croom, S. M., et al. 2016, A&A, 593, L8

McGaugh, S. S. 2018, Research Notes of the American Astronomical Society, 2, 156

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

McGreer, I. D., Mesinger, A., & Fan, X. 2011, MNRAS, 415, 3237

McQuinn, M. 2016, ARA&A, 54, 313

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

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

McWilliam, A., & Zoccali, M. 2010, ApJ, 724, 1491

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

Merrifield, M. R. 1992, AJ, 103, 1552

Mesinger, A. 2010, MNRAS, 407, 1328

Mesinger, A., & Furlanetto, S. 2007, ApJ, 669, 663

Mesinger, A., Furlanetto, S., & Cen, R. 2011, MNRAS, 411, 955

Miralda-Escudé, J., Cen, R., Ostriker, J. P., & Rauch, M. 1996, ApJ, 471, 582

Mitra, K., Chatterjee, S., DiPompeo, M. A., Myers, A. D., & Zheng, Z. 2018, MN-RAS, 477, 45

Miyamoto, M., & Nagai, R. 1975, PASJ, 27, 533

Mo, H., van den Bosch, F. C., & White, S. 2010, Galaxy Formation and Evolution

Mo, H. J., Mao, S., & White, S. D. M. 1998, MNRAS, 295, 319

Momjian, E., Carilli, C. L., Bañados, E., Walter, F., & Venemans, B. P. 2018, ApJ, 861, 86

Moore, B., Quinn, T., Governato, F., Stadel, J., & Lake, G. 1999, MNRAS, 310, 1147

Mortlock, D. 2016, 423, 187

Mortlock, D. J., Warren, S. J., Venemans, B. P., et al. 2011, Nature, 474, 616

Mroz, P., Udalski, A., Skowron, D. M., et al. 2018, ArXiv e-prints, arXiv:1810.02131

Naab, T., & Ostriker, J. P. 2017, ARA&A, 55, 59

Nataf, D. M., Udalski, A., Gould, A., Fouqué, P., & Stanek, K. Z. 2010, ApJL, 721, L28

Navarro, J. F., Frenk, C. S., & White, S. D. M. 1997, ApJ, 490, 493

Ness, M., Hogg, D. W., Rix, H.-W., Ho, A. Y. Q., & Zasowski, G. 2015, ApJ, 808, 16

Ness, M., & Lang, D. 2016, AJ, 152, 14

Nidever, D. L., Bovy, J., Bird, J. C., et al. 2014, ApJ, 796, 38

Novak, G. S., Ostriker, J. P., & Ciotti, L. 2011, ApJ, 737, 26

Oñorbe, J., Davies, F. B., Lukić, Z., Hennawi, J. F., & Sorini, D. 2018, arXiv e-prints, arXiv:1810.11683

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

Oñorbe, J., Hennawi, J. F., Lukić, Z., & Walther, M. 2017b, ApJ, 847, 63

Oh, S. P., & Furlanetto, S. R. 2005, ApJL, 620, L9

Omukai, K., Schneider, R., & Haiman, Z. 2008, ApJ, 686, 801

Oogi, T., Enoki, M., Ishiyama, T., et al. 2017, MNRAS, 471, L21

Oppenheimer, B. D., Segers, M., Schaye, J., Richings, A. J., & Crain, R. A. 2018, MNRAS, 474, 4740

Ortolani, S., Renzini, A., Gilmozzi, R., et al. 1995, Nature, 377, 701

Osmer, P. S. 1998, in Astronomical Society of the Pacific Conference Series, Vol. 146, The Young Universe: Galaxy Formation and Evolution at Intermediate and High Redshift, 1

Osmer, P. S. 2004, in Coevolution of Black Holes and Galaxies, ed. L. C. Ho, 324

Ostriker, J. P., Choi, E., Ciotti, L., Novak, G. S., & Proga, D. 2010, ApJ, 722, 642

Pâris, I., Petitjean, P., Rollinde, E., et al. 2011, A&A, 530, A50

Pâris, I., Petitjean, P., Aubourg, É., et al. 2018, A&A, 613, A51

Penzias, A. A., & Wilson, R. W. 1965, ApJ, 142, 419

Perlmutter, S., Aldering, G., Goldhaber, G., et al. 1999, ApJ, 517, 565

Piffl, T., Scannapieco, C., Binney, J., et al. 2014, A&A, 562, A91

Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2018a, arXiv e-prints, arXiv:1807.06209

Planck Collaboration, Akrami, Y., Arroja, F., et al. 2018b, arXiv e-prints, arXiv:1807.06211

Plummer, H. C. 1911, MNRAS, 71, 460

Polletta, M., Weedman, D., Hönig, S., et al. 2008, ApJ, 675, 960

Pont, F., Queloz, D., Bratschi, P., & Mayor, M. 1997, A&A, 318, 416

Portail, M., Gerhard, O., Wegg, C., & Ness, M. 2017, MNRAS, 465, 1621

Pouliasis, E., Di Matteo, P., & Haywood, M. 2017, A&A, 598, A66

Prochaska, J. X. 2017, Astronomy and Computing, 19, 27

Prochaska, J. X., Hennawi, J. F., & Herbert-Fort, S. 2008, ApJ, 675, 1002

Prochaska, J. X., O'Meara, J. M., & Worseck, G. 2010, ApJ, 718, 392

Prochaska, J. X., Hennawi, J. F., Lee, K.-G., et al. 2013, ApJ, 776, 136

Purcell, C. W., Bullock, J. S., Tollerud, E. J., Rocha, M., & Chakrabarti, S. 2011, Nature, 477, 301

Ramos Almeida, C., & Ricci, C. 2017, Nature Astronomy, 1, 679

Reed, S. L., McMahon, R. G., Martini, P., et al. 2017, MNRAS, 468, 4702

Reid, M. J., & Brunthaler, A. 2004, ApJ, 616, 872

Reid, M. J., Menten, K. M., Brunthaler, A., et al. 2014, ApJ, 783, 130

Rezaei Kh., S., Bailer-Jones, C. A. L., Hogg, D. W., & Schultheis, M. 2018, A&A, 618, A168

Richards, G. T., Vanden Berk, D. E., Reichard, T. A., et al. 2002, AJ, 124, 1

Riess, A. G., Filippenko, A. V., Challis, P., et al. 1998, AJ, 116, 1009

Robertson, B., Bullock, J. S., Cox, T. J., et al. 2006, ApJ, 645, 986

Rubin, V. C., Ford, Jr., W. K., & Thonnard, N. 1978, ApJL, 225, L107

Rubin, V. C., Ford, Jr., W. K., & Thonnard, N. 1980, ApJ, 238, 471

Runnoe, J. C., Cales, S., Ruan, J. J., et al. 2016, MNRAS, 455, 1691

Ryan-Weber, E. V., Pettini, M., Madau, P., & Zych, B. J. 2009, MNRAS, 395, 1476

Salpeter, E. E. 1964, ApJ, 140, 796

Sanders, D. B., Phinney, E. S., Neugebauer, G., Soifer, B. T., & Matthews, K. 1989, ApJ, 347, 29

Sanders, D. B., Soifer, B. T., Elias, J. H., et al. 1988, ApJ, 325, 74

Sartori, L. F., Schawinski, K., Trakhtenbrot, B., et al. 2018, MNRAS, 476, L34

Sartori, L. F., Schawinski, K., Koss, M., et al. 2016, MNRAS, 457, 3629

Schawinski, K., Koss, M., Berney, S., & Sartori, L. F. 2015, MNRAS, 451, 2517

Schlaufman, K. C., Rockosi, C. M., Allende Prieto, C., et al. 2009, ApJ, 703, 2177

Schmidt, M. 1963, Nature, 197, 1040

Schmidt, T. M., Hennawi, J. F., Worseck, G., et al. 2018, ApJ, 861, 122

Schneider, S. E., & Terzian, Y. 1983, ApJL, 274, L61

Schödel, R., Merritt, D., & Eckart, A. 2009, A&A, 502, 91

Selsing, J., Fynbo, J. P. U., Christensen, L., & Krogager, J.-K. 2016, A&A, 585, A87

Shang, Z., Brotherton, M. S., Wills, B. J., et al. 2011, ApJS, 196, 2

Sheinis, A. I., Bolte, M., Epps, H. W., et al. 2002, PASP, 114, 851

Shemmer, O., Brandt, W. N., Anderson, S. F., et al. 2009, ApJ, 696, 580

Shemmer, O., Trakhtenbrot, B., Anderson, S. F., et al. 2010, ApJL, 722, L152

Shen, Y., Greene, J. E., Strauss, M. A., Richards, G. T., & Schneider, D. P. 2008, ApJ, 680, 169

Shen, Y., & Ménard, B. 2012, ApJ, 748, 131

Shen, Y., Strauss, M. A., Oguri, M., et al. 2007, AJ, 133, 2222

Shen, Y., Strauss, M. A., Ross, N. P., et al. 2009, ApJ, 697, 1656

Shen, Y., Richards, G. T., Strauss, M. A., et al. 2011, ApJS, 194, 45

Shen, Y., Brandt, W. N., Richards, G. T., et al. 2016, ApJ, 831, 7

Shull, J. M., Stevans, M., & Danforth, C. W. 2012, ApJ, 752, 162

Silk, J. 1968, ApJ, 151, 459

Simcoe, R. A., Sullivan, P. W., Cooksey, K. L., et al. 2012, Nature, 492, 79

Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, AJ, 131, 1163

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

Sofue, Y., Honma, M., & Omodaka, T. 2009, PASJ, 61, 227

Sofue, Y., Tutui, Y., Honma, M., et al. 1999, ApJ, 523, 136

Soltan, A. 1982, MNRAS, 200, 115

Somerville, R. S., & Davé, R. 2015, ARA&A, 53, 51

Songaila, A. 2004, AJ, 127, 2598

Songaila, A., & Cowie, L. L. 2010, ApJ, 721, 1448

Springel, V., Frenk, C. S., & White, S. D. M. 2006, Nature, 440, 1137

Strömgren, B. 1939, ApJ, 89, 526

Tang, J.-J., Goto, T., Ohyama, Y., et al. 2017, MNRAS, 466, 4568

Tepper-García, T. 2006, MNRAS, 369, 2025

Toomre, A. 1977, in Evolution of Galaxies and Stellar Populations, ed. B. M. Tinsley
    & R. B. G. Larson, D. Campbell, 401

Trick, W. H., Coronado, J., & Rix, H.-W. 2018, arXiv e-prints, arXiv:1805.03653

Tully, R. B., & Fisher, J. R. 1977, A&A, 54, 661

Tytler, D., & Fan, X.-M. 1992, ApJS, 79, 1

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

Urry, C. M., & Padovani, P. 1995, PASP, 107, 803

Valiante, R., Schneider, R., Volonteri, M., & Omukai, K. 2016, MNRAS, 457, 3356

Vanden Berk, D. E., Richards, G. T., Bauer, A., et al. 2001, AJ, 122, 549

Venemans, B. P., Walter, F., Zschaechner, L., et al. 2016, ApJ, 816, 37

Venemans, B. P., McMahon, R. G., Walter, F., et al. 2012, ApJL, 751, L25

Venemans, B. P., Findlay, J. R., Sutherland, W. J., et al. 2013, ApJ, 779, 24

Venemans, B. P., Bañados, E., Decarli, R., et al. 2015, ApJL, 801, L11

Vestergaard, M., & Osmer, P. S. 2009, ApJ, 699, 800

Vestergaard, M., & Wilkes, B. J. 2001, ApJS, 134, 1

Visbal, E., Haiman, Z., & Bryan, G. L. 2014, MNRAS, 445, 1056

Volonteri, M. 2010, A&A Rv, 18, 279

Volonteri, M. 2012, Science, 337, 544

Walter, F., Riechers, D., Cox, P., et al. 2009, Nature, 457, 699

Wang, F., Wu, X.-B., Fan, X., et al. 2015, ApJL, 807, L9

Wang, F., Yang, J., Fan, X., et al. 2018, arXiv e-prints, arXiv:1810.11926

Wang, R., Carilli, C. L., Neri, R., et al. 2010, ApJ, 714, 699

Wang, R., Wagg, J., Carilli, C. L., et al. 2013, ApJ, 773, 44

Wang, R., Wu, X.-B., Neri, R., et al. 2016, ApJ, 830, 53

Watkins, L. L., van der Marel, R. P., Sohn, S. T., & Evans, N. W. 2018, ArXiv e-prints, arXiv:1804.11348

Wegg, C., & Gerhard, O. 2013, MNRAS, 435, 1874

Wegg, C., Gerhard, O., & Bieth, M. 2018, ArXiv e-prints, arXiv:1806.09635

Wegg, C., Gerhard, O., & Portail, M. 2015, MNRAS, 450, 4050

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, 157–169

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., Martini, P., & Cohn, J. D. 2008, MNRAS, 390, 1179

White, M., Myers, A. D., Ross, N. P., et al. 2012, MNRAS, 424, 933

White, S. D. M., & Rees, M. J. 1978, MNRAS, 183, 341

Wiese, W. L., & Fuhr, J. R. 2009, Journal of Physical and Chemical Reference Data, 38, 565

Willott, C. J., Bergeron, J., & Omont, A. 2015, ApJ, 801, 123

Willott, C. J., Omont, A., & Bergeron, J. 2013, ApJ, 770, 13

Willott, C. J., Delorme, P., Omont, A., et al. 2007, AJ, 134, 2435

Willott, C. J., Delorme, P., Reylé, C., et al. 2009, AJ, 137, 3541

Willott, C. J., Albert, L., Arzoumanian, D., et al. 2010, AJ, 140, 546

Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., et al. 2010, AJ, 140, 1868

Wu, X.-B., Wang, F., Fan, X., et al. 2015, Nature, 518, 512

Wyithe, J. S. B., & Bolton, J. S. 2011, MNRAS, 412, 1926

Wyithe, J. S. B., Loeb, A., & Carilli, C. 2005, ApJ, 628, 575

Xue, X.-X., Rix, H.-W., Ma, Z., et al. 2015, ApJ, 809, 144

Xue, X.-X., Rix, H.-W., & Zhao, G. 2009, Research in Astronomy and Astrophysics, 9, 1230

Yu, Q., & Tremaine, S. 2002, MNRAS, 335, 965

Zaldarriaga, M., Scoccimarro, R., & Hui, L. 2003, ApJ, 590, 1

Zel'dovich, Y. B. 1970, A&A, 5, 84

Zhang, L., Rix, H.-W., van de Ven, G., et al. 2013, ApJ, 772, 108

Zubovas, K., King, A. R., & Nayakshin, S. 2011, MNRAS, 415, L21

Zwicky, F. 1933, Helvetica Physica Acta, 6, 110

Acknowledgements

First and foremost, a huge thank you goes to my advisors, Joe, Hans-Walter, and David Hogg. Your mentorship, encouragement, and advice have been invaluable over the last few years, and the unfading enthusiasm and belief with which you have approached our projects has been truly inspiring.

To Fabian, thank you for always providing calm and kind support, advice, and feedback, no matter what problems or questions I had.

To Fred, thank you for patiently answering all my many questions over the last few years.

A big thank you should also go to Christian Fendt for providing so many opportunities through the IMPRS school, as well as for agreeing to referee my thesis.

A special thanks goes to all my fellow students at MPIA for providing a truly enjoyable working atmosphere, great friendships, and innumerable fun lunch and coffee breaks. To Sara, thank you for showing me your beautiful home country and for providing me with Persian sweets, fresh pistachios, and Farsi lessons – kheili mamnoon, golam. Michi, thanks for always being up for a chat and for your continuous `matplotlib` and IT support. To Wilma, thank you for your friendship and for the countless discussions about color schemes and acronyms. Tobias, thanks for organizing the IMPRS retreat with me and for just being a great friend. Johannes, thank you for the many delightful rides up to MPIA. To Chiara, thanks for your moral support and for always providing Italian cookies.

A particular thanks also goes to Fred, Michi, Johannes, Rich, Wilma, Tobias, and Chiara, for proof-reading various parts of my thesis.

To all the Italians (and nearly-Italians), grazie mille for providing laughs and coffee early every morning!

Rich, thank you for always making me smile, for your love and your never ending patience. I am excited to start our new adventure together with you in Boston.

Finally, my family. Mama, Papa, Caro, Oma: I am forever thankful for your unparalleled love and support, as well as for the continuous supply of Spätzle, Maultaschen, and Rosinenbrötchen. Thank you for everything.