Probing the connection between the intergalactic medium and galaxies with quasar absorption-line spectroscopy

Charles William Finn

A Thesis presented for the degree of
Doctor of Philosophy

Extragalactic Astronomy and Cosmology Group
Institute for Computational Cosmology
Department of Physics
University of Durham
UK

April 2015

Dedicated to

My friends and family

Probing the connection between the intergalactic medium and galaxies with quasar absorption-line spectroscopy

Charles W. Finn

Submitted for the degree of Doctor of Philosophy
April 2015

Abstract

In this thesis, we examine the relationship between the metal-enriched intergalactic medium (IGM) and galaxies at $z < 1$. In particular, we investigate the nature and consequence of feedback from active galactic nuclei (AGN) and supernovae, which shape the evolution of galaxies and are responsible for enriching the IGM with metals. The IGM is surveyed in ultraviolet (UV) absorption lines against background quasars (QSOs), whilst galaxies are surveyed in emission by means of optical photometry and spectroscopy. Simulated samples of IGM absorption systems and galaxies are also extracted from the Evolution and Assembly of GaLaxies and their Environments (EAGLE) cosmological hydrodynamical simulation for critical comparison with the data. We present the results of two primary studies that are designed to address key questions on the nature and consequence of feedback:

1. We examine complex absorption profiles in the spectrum of a QSO at $z \sim 1$, that trace a metal-rich outflow originating from the host galaxy. We show that these absorption profiles originate from dense, sub-pc scale gas clumps at distances of a few kpc from the central AGN. The gas is likely to be dynamically unstable, and is potentially far from ionization equilibrium. We favour a scenario in which the clumps are formed in-situ, and are entrained in a hot ($T > 10^6$ K) outflowing wind that may trace the majority of the mass, but is undetected in the UV. These observations provide a detailed

set of constraints on the nature of feedback in QSO host galaxies.

2. We investigate the distribution and dynamics of metal-enriched gas around galaxies at $z < 1$ through the two-point cross- and auto-correlation functions of O VI absorbers and galaxies. We find that O VI absorbers show little velocity dispersion with respect to galaxies on $\sim$ Mpc scales ($\lesssim$ 100 km s^{-1}). O VI absorbers and galaxies may not linearly trace the same underlying distribution of matter in general. In particular, the distribution of O VI around galaxies could be more extended than the distribution of galaxies around themselves. As a result, we speculate that a fraction of the O VI absorbers might trace the warm-hot intergalactic medium (WHIM). In addition, we find that O VI absorbers are equally likely to inhabit the same regions typically occupied by star-forming galaxies as they are to inhabit the same regions typically occupied by non star-forming galaxies. Furthermore, O VI absorbers are either not ubiquitous to galaxies, or their distribution around them is patchy on $\gtrsim$ 100 kpc scales (or both). Comparisons with the EAGLE simulation indicate an unprecedented level of agreement between both the anisotropies in the cross-correlation functions, and their amplitudes and slopes. This result is both remarkable and robust, since the feedback parameters in EAGLE are not adjusted to ensure a match with observations such as these. We therefore find that the feedback implementations currently employed in cosmological hydrodynamical simulations are sufficient to provide meaningful insights on the dynamics and distribution of the metal-enriched gas surrounding galaxies.

Declaration

The work in this thesis is based on research carried out at by the author at the Extragalactic Astronomy and Cosmology Group and the Institute for Computational Cosmology, Department of Physics, Durham University, UK, under the supervision of Prof. Simon L. Morris and Dr Tom Theuns. No part of this thesis has been submitted elsewhere for any other degree or qualification.

The results and analysis presented in Chapter 3 have been published in the following paper:

- *A compact, metal-rich, kpc-scale outflow in FBQS J0209−0438: detailed diagnostics from HST/COS extreme UV observations*, Finn, C. W.; Morris, S. L.; Crighton, N. H. M.; Hamann, F.; Done, C.; Theuns, T.; Fumagalli, M.; Tejos, N.; Worseck, G., MNRAS 440, 3317–3340 (2014).

The results and analysis presented in Chapter 6 are in preparation for publication. During this time, the author has also contributed to the following published work:

- *Large scale structure in absorption: gas within and around galaxy voids*, Tejos, N.; Morris, S. L.; Crighton, N. H. M.; Theuns, T.; Altay, G.; Finn, C. W., MNRAS 425, 245–260 (2012).

- *On the connection between the intergalactic medium and galaxies: The H* I*–galaxy cross-correlation at* $z \lesssim 1$, Tejos, N.; Morris, S. L.; Finn, C. W.; Crighton, N. H. M.; Bechtold, J.; Jannuzi, B. T.; Schaye, J.; Theuns, T.; Altay, G.; Le Fèvre, O.; Ryan-Weber, E.; Davé, R., MNRAS 437, 2017–2075 (2014).

Most of the work in this thesis is entirely that of the author unless referenced to the contrary in the text or in the following list:

- Acquisition and reduction of the infrared spectroscopy of FBQS J0209−0238 from Magellan/FIRE, presented in Chapter 3, was performed by Michele Fumagalli.
- Galaxy data and spectral line measurements from the GAMA survey that are not part of the current public data release, presented in Chapter 4, were obtained from Madusha Gunawardhana, Peder Norberg, and the GAMA survey team.
- The extraction of synthetic QSO sight-lines through the EAGLE simulation with SPECWIZARD, described in Chapter 5, was undertaken by Robert Perry.

Acknowledgements

Firstly, I'm am deeply grateful to my parents, Duncan and Edwina, and my two younger sisters, Georgina and Rachel, for their unwavering love and support throughout my life and during this important time.

Many thanks to must go to Simon Morris, who has been an excellent supervisor, and provided me with a wealth of support and guidance. He has been instrumental in my development as an independent working scientist, encouraging me to question, to think critically, and to give due consideration to all manner of interpretations in light of new results. I also thank my second supervisor, Tom Theuns, for his insight, and for always coming up with interesting new ideas to explore. Special thanks must also go to Neil Crighton, who has been a constant source of help and support, and is hugely responsible for my development as a competant computer programmer. Much gratitude also goes to Nicolas Tejos, who was a major source of inspiration to me early in my PhD, and has been a terrific collaborator and friend. Thanks also to Rich Bielby and Michele Fumagalli for your enthusiasm, help and insight.

Life in Durham wouldn't have been the same without all the wonderful friends I have made here. I would like to thank my office mates, past and present, all those who I worked with and participated in journal club meetings with; thank you for your help and ideas, and for creating such a nice, friendly working atmosphere. Special thanks to: Tamsyn, Rachel, Peter M., Peter C., Emma, Ewan, Will, Helen, Alice, Chris, Violeta, Madusha, Michele, Gabriel, Tim, Steph, Nicolas, Rich, Alasdair, Danny, Matthieu, Ben, Andrew, Paddy, Sownak, Alex, Flora, Ben, Alice, Alix, George, Nikki, James C., James T., James S., Rob, Wojciech, Mike and Michelle!

Contents

List of Figures

List of Tables

Chapter 1 | *Introduction*

The aim of this thesis is to investigate the relationship between galaxies and the metal-enriched intergalactic medium (IGM). In this chapter, we will define what we mean when referring to these concepts, and summarise progress in the literature towards our understanding in this area. But first, to set the scene, we shall briefly review the observable facts and founding theoretical principles behind the consensus galaxy formation paradigm, and describe the key observables.

1.1 The ΛCDM paradigm

The consensus model of the Universe we inhabit is one predominantly formed of exotic forms of matter and energy entirely different to those found on earth in our every day experience. These completely hypothetical entities are referred to as *dark matter* and *dark energy*, and neither have been directly confirmed by experiment. Astonishing as this may sound, there are nevertheless compelling arguments for their existence.

Dark matter is so-named because it is hypothesised to interact only through gravity and the electroweak force, and therefore does not emit or absorb any form of electromagnetic radiation. The presence of this invisible form of matter has been inferred from a number of observations, notably from the velocity dispersions of galaxies in clusters (e.g. Zwicky, 1937), the rotation curves of spiral galaxies (e.g. Rubin & Ford, 1970; Rubin et al., 1980), the gravitational lensing induced by massive galaxy clusters (e.g. Clowe et al., 2006), and the power spectrum of the cosmic microwave background (CMB) (e.g. Komatsu et al., 2011; Planck Collaboration et al., 2014). The latest measurements from the latter suggest that nearly 85% of all the matter in the Universe is dark matter. The small amount remaining is that which can, in principle, be observed directly via the emission and/or

absorption of electromagnetic radiation, and is referred to as 'baryonic' matter.[1]

The other major component to the Universe is dark energy. Nothing is known of this elusive component, except that it must be responsible for the accelerated expansion of the Universe. A potted history behind this is as follows. Early observations of local galaxies indicated that their velocities are proportional to their distances (e.g. Hubble, 1929), and it was quickly realised that this may be naturally explained by a set of solutions to Einstein's General Relativity (GR) equations (Einstein, 1916) that predict a largely isotropic and homogeneous Universe must be expanding at some point in its history (e.g. Friedmann, 1922; Lemaître, 1927; Robertson, 1935; Walker, 1935). An extrapolation of these models also points to a time in the past when all the components of the Universe were concentrated into a single point of zero size and infinite energy density - a *singularity*. This idea has developed into a theory for the origin of the Universe, popularly termed the 'Big Bang' theory. Observational evidence for such a scenario comes from measurements of the CMB (e.g. Penzias & Wilson, 1965), interpreted as the left-over radiation from the 'surface of last scattering' (the point in time after the Big Bang when radiation and matter decoupled, e.g. Dicke et al., 1965), and the primordial abundance of light elements (e.g. Pettini & Cooke, 2012), which are consistent with those predicted from Big Bang nucleosynthesis (e.g. Alpher et al., 1948).

Determining the time evolution of the Universe requires some knowledge of its matter-energy density, which determines its geometrical curvature. Observations of the CMB favour a Universe with effectively zero curvature, and for such a Universe that is matter dominated, the gravitational force acts to decelerate the expansion. However, observations of type Ia supernovae in distant galaxies have shown that the expansion of the Universe is *accelerating* (e.g. Riess et al., 1998; Perlmutter et al., 1999). This behaviour requires that there be an additional component to the energy density of the Universe that opposes the force of gravity. It is this additional component that is referred to as dark energy, which can be described

[1]Electrons and neutrinos (fermions) are usually included in the definition of this component, although they are considerably lighter than baryons and therefore do not contribute significantly to the total matter density of the Universe.

by the addition of a constant to the field equations of GR, known as the *cosmological constant*, Λ. This must be present even in the absence of matter, and can therefore be thought of as an intrinsic property of the vacuum. The energy associated with the vacuum is in fact predicted by quantum field theory (e.g. Weinberg, 1989; Carroll, 2001), but differs from the value inferred from astronomical data by ~ 120 orders of magnitude! Reconciling the tension between measurements of Λ, and those predicted in theory, will likely revolutionise our understanding of the Universe.

The cosmological paradigm we assume throughout this thesis is referred to as the ΛCDM paradigm, to mean that we exist in a Universe whose evolution is primarily dictated by the presence of dark energy in the form of a cosmological constant, Λ, and a form of dark matter known as *cold dark matter*. By 'cold' we mean that the dark matter particle is both massive and non-relativistic. Such a dark matter particle is needed to reconcile the measured clustering of galaxies with the hypothesis that they are related to small over-densities at early times as are traced by the (extremely) small, but statistically significant temperature fluctuations in the CMB. Since dark matter should be collisionless and dissipationless (it interacts only via gravity and the electroweak force), primordial baryonic material is allowed to radiatively cool, condense, and form stars (and galaxies) at the peaks of the underlying dark matter distribution (e.g. Press & Schechter, 1974; Rees & Ostriker, 1977; White & Rees, 1978; White & Frenk, 1991; Kereš et al., 2005) in so-called dark matter 'haloes'. This idea forms the basis for galaxy formation under a ΛCDM paradigm, and is a major backdrop to the work in this thesis. For a thorough and extended description, we refer the reader to Mo et al. (2010).

1.2 Cosmological principles

Throughout this work, we assume a ΛCDM cosmology with best-fit parameters as obtained from the latest measurement of the CMB power spectrum (Planck Collaboration et al., 2014) unless stated otherwise. Here we describe a few key aspects of this model, as are frequently adopted throughout the literature.

1.2.1 Geometry and dynamics of the expanding Universe

For an isotropic and homogeneous Universe, the most general space-time metric can be written as follows:

$$\mathrm{d}s^2 = c^2\mathrm{d}t^2 - R^2(t)\left(\frac{\mathrm{d}r^2}{1-kr^2} + r^2\mathrm{d}\theta^2 + r^2\sin^2(\theta)\mathrm{d}\phi^2\right), \tag{1.2.1}$$

where c is the speed of light, t is cosmological time, r, θ, and ϕ are spherical coordinates (radial distance, polar angle and azimuthal angle respectively), k denotes the geometrical curvature of the Universe ($k \in \{-1, 0, 1\}$, corresponding to negative, zero, and positive curvature respectively), and $R(t)$ is the time-dependent scale factor, which accounts for the expansion of the Universe, making r, θ, and ϕ time-independent *comoving* coordinates. It is common to define a dimensionless scale factor as

$$a(t) \equiv \frac{R(t)}{R_0}, \tag{1.2.2}$$

where R_0 is the present value of the scale-factor, so that $a = 1$ at the present time.

For a metric such as this, it is easy to see how we may recover a mathematical description for Hubble's observations of an expanding Universe, i.e. Hubble's law, $v = Hr$, where v is the recession velocity of a receding galaxy, and H is Hubble's constant. At small separations, where things look Euclidean, the proper separation of two objects is just $R(t)\mathrm{d}r$, so we obtain Hubble's law as

$$H = \frac{\dot{R}}{R}. \tag{1.2.3}$$

In practice, we can measure this *radial* recession velocity in terms of a *redshift*, z, which describes the shift in spectral lines (see Section 1.4), although this is only valid for small separations, since at large separations spatial curvature becomes important. In general, since photons travel on null geodesics,[2] it is clear from the

[2]Geodesics with zero proper time, i.e. $\mathrm{d}s \equiv c^2\mathrm{d}\tau^2 = 0$.

metric that

$$r = \int \frac{c\mathrm{d}t}{R(t)}. \tag{1.2.4}$$

The comoving distance remains constant with time by definition, and the integral in time spans from t_{emit} to t_{obs}, which are the times of emission and observation of a photon. Photons emitted at later times are received at later times, but r must nevertheless remain constant, which enforces the condition $\mathrm{d}t_{\mathrm{emit}}/\mathrm{d}t_{\mathrm{obs}} = R(t_{\mathrm{emit}})/R(t_{\mathrm{obs}})$. This means that events from distant galaxies time dilate according to how much the Universe has expanded since the photons were emitted. It is easy to see that this can be expressed in terms of frequency, ν, and thus

$$\frac{\nu_{\mathrm{emit}}}{\nu_{\mathrm{obs}}} \equiv 1 + z = \frac{R(t_{\mathrm{obs}})}{R(t_{\mathrm{emit}})}. \tag{1.2.5}$$

In terms of the dimensionless scale factor, we have that $a(t) = (1 + z)^{-1}$, so it is clear that photon wavelengths stretch with the Universe.

The equation of motion for the scale factor is required if we are to know the cosmological time at which photons are emitted (or absorbed) by any particular extragalactic object. With equation (1.2.1) in combination with Einstein's field equations, we can obtain the so-called Friedmann equation:

$$\dot{R}^2 - \frac{8\pi G}{3}\rho R^2 = -kc^2. \tag{1.2.6}$$

It is important to note that this equation applies for all contributions to ρ, i.e. those from matter, radiation, and the vacuum, so it is independent of the equation of state. One astonishing feature is that there is a direct connection between the density of the Universe and its geometry. This is seen by considering that there is a *critical density*, ρ_{crit}, that will yield $k = 0$, making the comoving part of the metric in equation (1.2.1) look Euclidean:

$$\rho_{\mathrm{crit}} = \frac{3H^2}{8\pi G}. \tag{1.2.7}$$

If $\rho = \rho_{\text{crit}}$, the Universe looks Euclidean, and is spatially *flat*.[3] On the other hand, if $\rho > \rho_{\text{crit}}$, the Universe is spatially *closed*, and if $\rho < \rho_{\text{crit}}$, the Universe is spatially *open*. Upon solving the Friedmann equation, it can be seen that closed Universes are bound, and open Universes are unbound, to mean that the Universe either re-collapses after a period of expansion, or continues to expand forever. Without explicitly solving the Friedmann equation, the case of the latter can easily be seen as follows. First, we note that any open model will have an undecelerated expansion so long as its equation of state is such that ρR^2 is a declining function of R. In this case, the potential energy (second term on the left-hand-side of equation (1.2.6)) becomes negligible with the total (right-hand-side of equation (1.2.6)), and the kinetic term, $\dot{R}^2$, tends to c such that $R = ct$. In this case, all objects are moving with constant velocity, and the Universe continues to expand forever. To prove the case of a closed Universe requires explicitly solving the Friedmann equation, which we shall not do here.

It is common to define a dimensionless *density parameter* as the ratio of the density to the critical density:

$$\Omega \equiv \frac{\rho}{\rho_{\text{crit}}} = \frac{8\pi G\rho}{3H^2}. \tag{1.2.8}$$

We can then re-write the Friedmann equation as

$$\frac{kc^2}{H^2R^2} = \Omega - 1. \tag{1.2.9}$$

A flat ($k = 0$) Universe therefore requires $\Omega = 1$.

Since R, H, and Ω change with time, we denote their present day values with a zero subscript. From the Friedmann equation, we can see that the present day value for the scale factor is

$$R_0 = \frac{c}{H_0}\left(\frac{\Omega_0 - 1}{k}\right)^{-\frac{1}{2}}. \tag{1.2.10}$$

[3]Note that we still have curved spacetime in a flat Universe.

This value becomes infinitely large as Ω_0 approaches unity, and so any model with Ω_0 very close to unity is virtually indistinguishable from a $k = 0$ model. In practice, we shall denote Ω_0 as Ω, and refer to the density parameter at all other epochs by writing it as a function of redshift, $\Omega(z)$.

We also define here the commonly used dimensionless (current) Hubble parameter as

$$h \equiv \frac{H_0}{100 \text{ km s}^{-1} \text{ Mpc}^{-1}}. \tag{1.2.11}$$

In these terms, the current density of the Universe is

$$\begin{aligned} \rho_0 &= 1.88 \times 10^{-26} \Omega h^2 \text{ kg m}^{-3} \\ &= 2.78 \times 10^{11} \Omega h^2 M_\odot \text{ Mpc}^{-3}, \end{aligned} \tag{1.2.12}$$

and the current value of the scale factor is

$$R_0 = 3000 \, |\Omega - 1|^{-\frac{1}{2}} h^{-1} \text{ Mpc}. \tag{1.2.13}$$

So far, we have not considered the effect of pressure on the evolution of the Universe, which is important even for a matter dominated Universe at early times when the temperature and density is high. To consider the effects of pressure, we'll take the time derivative of the Friedmann equation. We can deal with the time derivative of the density by invoking conservation of energy: $\mathrm{d}(\rho c^2 R^3) = -p\mathrm{d}(R^3)$, so that

$$\ddot{R} = -\frac{4\pi G R}{3} \left(\rho + \frac{3p}{c^2} \right). \tag{1.2.14}$$

The unusual element to this equation is the occurrence of the term $\rho + 3p/c^2$, which in fact arises because Einstein's gravitational field equations in the weak-field limit yield

$$\nabla^2 \Phi = 4\pi G \left(\rho + \frac{3p}{c^2} \right). \tag{1.2.15}$$

Pressure, therefore, acts as an extra form of gravity, which is outside the realms of our everyday experience, but turns out to be important in cosmology.

Returning to the Friedmann equation in the form of equation (1.2.9), we recall that Ω includes all contributions to the density of the Universe, from both matter and energy. Matter and radiation ($\Omega_{\rm m}$ and $\Omega_{\rm r}$) are obvious, but the third, Ω_v, corresponding to the energy density of the vacuum, is less so. For this we first note that the gravitational potential can be reduced to that which solves Poisson's equation: $\nabla^2\Phi = 4\pi G\rho$. In 1917, Einstein argued that, by symmetry, in a Universe with constant density ρ, the gravitational potential, Φ, must also be constant, but this does not solve Poisson's equation. To allow for this, its modification is as follows:

$$\nabla^2\Phi + \Lambda\Phi = 4\pi G\rho, \tag{1.2.16}$$

where Λ is a new constant of nature, termed the *cosmological constant*. We can re-write this with a new quantity, $\rho_{\rm rep} = \Lambda\Phi/4\pi G$, as follows:

$$\nabla^2\Phi = 4\pi G(\rho - \rho_{\rm rep}), \tag{1.2.17}$$

i.e. with a constant *repulsive* density that has *antigravity* properties. This must be present even in the absence of matter, and is therefore interpreted as an intrinsic property of the vacuum. Weird as this sounds, for a vacuum with non-zero energy density, it can easily be shown that the vacuum must also have a non-zero pressure, with a negative-pressure equation of state:

$$p_{\rm vac} = -\rho_{\rm vac}c^2. \tag{1.2.18}$$

By analogy with equation (1.2.15), this indeed acts against gravity to cause a large-scale repulsion. From a cosmological perspective, this is then the theoretical explanation for what we term *dark energy*.

In order to solve the Friedmann equation for the expansion history of the Universe, we need to know how the density changes as R changes, which we can

achieve by dividing the contents of the Universe into pressureless matter ($\rho \propto R^{-3}$), radiation ($\rho \propto R^{-4}$), and vacuum energy (ρ constant). In terms of quantities that may be obtained from astronomical observations, we then have

$$\frac{8\pi G\rho}{3} = H_0^2 \left(\Omega_\mathrm{v} + \Omega_\mathrm{m} a^{-3} + \Omega_\mathrm{r} a^{-4}\right). \tag{1.2.19}$$

Using this, we can then always integrate the Friedmann equation numerically to find $R(t)$.

1.2.2 Observations under a ΛCDM framework

At many points during this thesis, we shall use the observed redshift, z, of an extragalactic object to determine its comoving distance. From equation (1.2.4) we have

$$\begin{aligned} r &= c \int_{t_\mathrm{emit}}^{t_\mathrm{obs}} \frac{\mathrm{d}t}{a(t)} \\ &= c \int_{a(t_\mathrm{emit})}^{a(t_\mathrm{obs})=1} \frac{\mathrm{d}a}{a(t)\dot{a}(t)}, \end{aligned} \tag{1.2.20}$$

which in terms of observable parameters is

$$r = \frac{c}{H_0} \int_0^z \mathrm{d}z \ \left(\Omega_\mathrm{r}(1+z)^4 + \Omega_\mathrm{m}(1+z)^3 + \Omega_\mathrm{v} + (1-\Omega)(1+z)^2\right)^{-\frac{1}{2}}. \tag{1.2.21}$$

The latest observational constraints for these parameters are $\Omega_\mathrm{r} \approx 0.000$, $\Omega_\mathrm{m} \approx 0.307$, $\Omega_\mathrm{v} \approx 0.693$, and $H_0 \approx 67.77$ km s^{-1} Mpc^{-1} (Planck Collaboration et al., 2014), with $\Omega \approx 1$ and $k \approx 0$. This corresponds to a Universe under a ΛCDM paradigm that is consistent with being spatially flat, and presently expanding at an increasing rate. It is of interest to note that only $\approx 16\%$ of the contribution to Ω_m comes from baryonic matter. The majority is *dark matter*, presumably 'cold'.

Another quantity of interest is the physical transverse size of an extragalactic object, or the physical transverse separation between two extragalactic objects. For the ($k = 0$) Universe implied by the cosmological parameters above, we may write

the space-time metric in equation (1.2.1) as

$$\mathrm{d}s^2 = c^2\mathrm{d}t^2 - R^2(t)\left(\mathrm{d}r^2 + r^2\mathrm{d}\psi^2\right), \tag{1.2.22}$$

where $\mathrm{d}\psi^2 \equiv \mathrm{d}\theta^2 + \sin^2(\theta)\mathrm{d}\phi^2$ corresponds to the angular size on the sky. The spatial part of this metric then tells us that we can find the proper transverse size element as

$$\mathrm{d}l_\perp = \mathrm{d}\psi R(z)r = \mathrm{d}\psi\frac{R_0 r}{1+z}. \tag{1.2.23}$$

The element of proper distance in the radial direction is $R(z)\mathrm{d}r$, so we can also calculate the cosmological volume element. For a region of sky whose area is A steradians, combined with a depth in comoving distance $\mathrm{d}r$, the proper volume element corresponds to

$$\mathrm{d}V = \left(R(z)r\mathrm{d}\psi\right)^2 \times R(z)\mathrm{d}r = AR(z)^3r^2\mathrm{d}r. \tag{1.2.24}$$

Often we are more interested in the comoving volume, which is as simple as replacing $R(z)$ with R_0 in the above equation.

One of the most important relationships for observational cosmology is that between monochromatic flux density and luminosity, which is encountered multiple times throughout this thesis. We start by assuming that an extragalactic object emits light isotropically, so that photons pass through a sphere surrounding the object with uniform flux density. We next consider the origin of the metric in equation (1.2.1) as being centred on the source, but because of homogeneity, the comoving distance between the source and the observer is the same as it would be had we placed the origin at the observer's location. The photons from the source therefore pass through a sphere of proper surface area $4\pi(R_0 r)^2$, on which the observer sits. However, redshift affects the flux density in four further ways:

1. Photon energies are redshifted, reducing the flux density by a factor of $(1+z)$.

2. Photon arrival rates are time dilated, reducing the flux density by another factor of $(1+z)$.

3. The bandwidth, $\mathrm{d}\nu$, is reduced by a factor of $(1+z)$, which increases the energy flux per unit bandwidth by one power of $(1+z)$.

4. The observed photons at frequency ν_0 were emitted at frequency $\nu_0(1+z)$.

Overall, the flux density is the luminosity at a frequency $\nu_0(1+z)$, divided by the total area, divided by $(1+z)$:

$$S_\nu(\nu_0) = \frac{L(\nu_0(1+z))}{4\pi R_0^2 r^2 (1+z)}. \tag{1.2.25}$$

Given an estimate of the functional form for $L(\nu)$, we can integrate over ν_0 to get the total flux, known as the *bolometric* flux.

Based on the form of the equations presented above, we can present two further conventionally used comoving distance measures:

$$d_\mathrm{A} \equiv \frac{r}{(1+z)} \tag{1.2.26}$$

and

$$d_\mathrm{L} \equiv (1+z)r, \tag{1.2.27}$$

which are the *angular diameter distance* and the *luminosity distance* respectively. We plot these distance measures in Figure 1.1, along with the comoving distance in dimensionless form as a function of redshift. It is interesting to note in the middle panel that the angular diameter distance turns over at $z \gtrsim 1.5$. This is due to the gravitational deflection of light, which produces a focusing effect, as if we were viewing distance objects through a massive fish-eye lens!

Finally, we define here the commonly used *absolute magnitude*, M, as the apparent magnitude, m, of an extragalactic source if it were observed at a distance of 10 pc. This is simply a measure of luminosity. Absolute magnitudes are affected by a shift of the emitted spectrum in frequency due to the redshift. This is accounted for by the so-called *K correction*, $K(z)$, which gives the difference between the observed dimming with redshift for a fixed waveband and that expected on

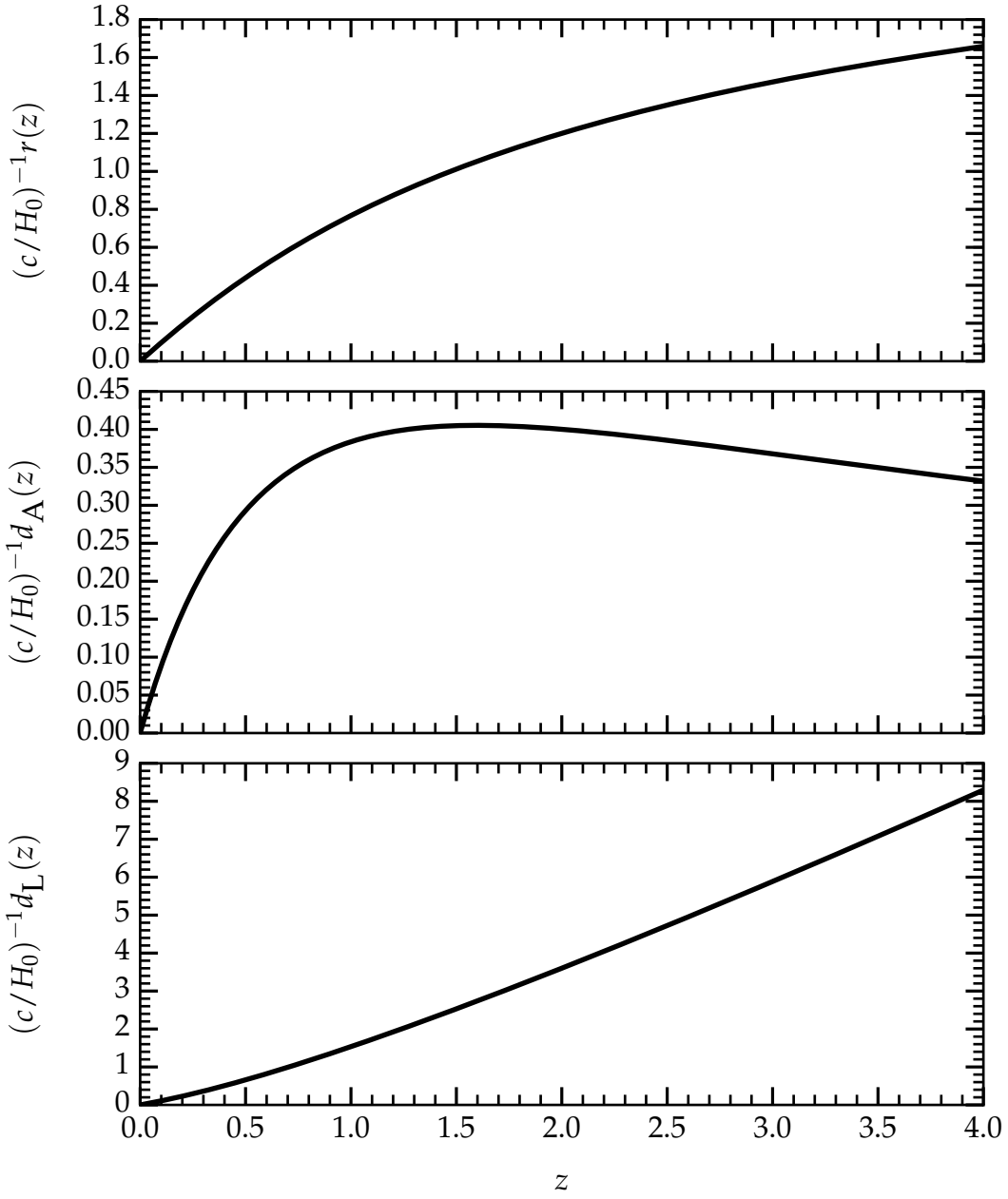

Figure 1.1: Dimensionless distance measures in cosmology as a function of redshift. From top to bottom: comoving distance, angular diameter distance and luminosity distance.

bolometric grounds. We then have

$$m \equiv M + 5\log_{10}\left(\frac{d_{\mathrm{L}}}{10\ \mathrm{pc}}\right) + K(z). \tag{1.2.28}$$

1.3 Galaxies and the intergalactic medium

The major topic of this thesis is the IGM and its relationship to galaxies. Here we define what we mean when referring to these concepts, and summarise some key findings in the literature.

Firstly, a 'galaxy' is a commonly used term in astronomy, but one that is not that well defined (see Forbes & Kroupa, 2011, for a recent discussion). Often, from a ΛCDM standpoint, galaxies are defined as being bound baryonic structures that have formed stars, residing within the gravitational potential of a dark matter halo. However, it would be nicer to find a definition that is not model dependent, i.e. one that does not rely upon the presence of (as of yet still hypothetical) dark matter haloes. One may therefore simply propose that a galaxy is a gravitationally bound system of stars and other baryonic components. However, there are a growing number of astronomical objects with extremely low luminosities and surface brightnesses that fill the gap between objects traditionally thought of as galaxies, and far smaller collections of stars known as 'star clusters' (a few $\times 10^2 M_\odot$; Misgeld & Hilker, 2011). To address this issue, Willman & Strader (2012) proposed a new definition, stating: '*a galaxy is a gravitationally bound collection of stars, whose properties cannot be explained by a combination of baryons and Newton's laws of gravity*', which is rooted in the idea that galaxies should show evidence for what is interpreted to be dark matter in the consensus theoretical paradigm. Nevertheless, there may be other kinds of objects to consider, such as the so-called 'dark galaxies', having cold gas but very few, or no stars (e.g. Cantalupo, 2010; Gnedin & Kravtsov, 2010; Kuhlen et al., 2012; Cantalupo et al., 2012). For the purposes of this thesis, we shall simply refer to a galaxy as a gravitationally bound system, whose observed spectrum is consistent with that of a population of stars and/or gas in optical emission, having an integrated luminosity greater than that typical of Milky Way globular clusters ($\gtrsim 10^{6-7} L_\odot$).

The IGM is loosely defined as the baryonic material that is not part of a galaxy. Since galaxies likely do not evolve in closed boxes - there is almost certainly a cycle of material flowing in and out (see Putman et al., 2012, for a recent review) - the

definition of the IGM is, inevitably, closely linked to that of a galaxy. It is tempting to define a boundary at which a galaxy ends, and the IGM begins, but there is no obvious choice for such a boundary. This was recently discussed at length by Shull (2014), who consider a number of different boundary conditions, including the commonly used 'virial radius', denoting the characteristic separation between collapsed structures in dynamical equilibrium, and external matter, which may be infalling due to gravity, or outflowing due to winds from supernovae explosions and/or active galactic nuclei (AGN). Characteristic radii such as this form a useful reference point that scales naturally with galaxy stellar and halo mass, however, there is no evidence or theoretical motivation for a sudden physical change in the properties of the IGM at this, or any other boundary one cares to draw around galaxies in general.

For the purposes of this thesis, we simply consider the IGM as baryonic material that is not in the vicinity of galaxy discs and spheroids, where the majority of the stars and sites of active star formation are to be found. Gas in these regions is typically referred to as the interstellar medium (ISM), although it is important to note that there is no clear definition for where the ISM transitions into the IGM either. For example, there is growing evidence for the existence of hot ($\sim 10^6$ K) gas 'coronae' in and around galaxies, which are bound and in hydrostatic equilibrium (e.g. Spitzer, 1956; Forman et al., 1985; Wang et al., 2001; Strickland et al., 2004a; Fukugita & Peebles, 2006; Li & Wang, 2013a), and may form due to both infalling gas that is shock heated to approximately the dark matter halo virial temperature[4] (e.g. Rees & Ostriker, 1977; White & Rees, 1978; White & Frenk, 1991; Crain et al., 2013; Li et al., 2014), and due to shocks associated with stellar winds and supernova explosions (e.g. Mathews, 1990; Ciotti et al., 1991; Strickland et al., 2004b; Parriott & Bregman, 2008; Li & Wang, 2013b). The full extent of the corona is uncertain, but it may also play host to the cooler (but frequently ionized) gas observed in absorption against background quasars (QSOs) (and against background stars in the Milky Way, e.g. Wakker & van Woerden, 1997; Sembach et al., 2003; Shull

[4] $T_{\rm vir} = 10^6 (v_{\rm circ}/167 \text{ km s}^{-1})^2$ K, where $v_{\rm circ}$ is the circular velocity. Gas at this temperature is in quasi-hydrostatic equilibrium with the dark matter halo.

et al., 2009; Lehner & Howk, 2011; Tumlinson et al., 2013; Werk et al., 2013, 2014) in a region often termed the circumgalactic medium (CGM), and it is almost certainly coincident with the observed 'extraplanar' Hα emission (e.g. Hoopes et al., 1999; Rossa & Dettmar, 2000, 2003a,b; Miller & Veilleux, 2003a,b; Strickland et al., 2004a) and neutral H I 21 cm emission (e.g. Chaves & Irwin, 2001; Fraternali et al., 2002; Matthews & Wood, 2003; Barbieri et al., 2005; Boomsma et al., 2005; Oosterloo et al., 2007). All of these cooler gas components may arise due to gas condensations in the corona (perhaps driven by supernova events; e.g. Bregman, 1980; Fraternali & Binney, 2008; Marinacci et al., 2010; Marasco et al., 2012; Fraternali et al., 2013), or from gas falling in at larger distances with potentially lower metallicity (e.g. Richter, 2012; Hernandez et al., 2013; Schmidt et al., 2014). In addition, analytic theory and cosmological hydrodynamical simulations indicate the existence of cold ($\sim 10^{4-5}$ K), metal-poor, filamentary gas flows that penetrate galaxy haloes without shock heating, predominantly in low mass haloes at high redshift, and potentially delivering gas from distances of several hundred kpc directly to the sites of star formation (e.g. Birnboim & Dekel, 2003; Kereš et al., 2005; Dekel & Birnboim, 2006; Kereš et al., 2009b; Dekel et al., 2009). The prevalence of these filamentary gas flows is somewhat debated (e.g. Sijacki et al., 2012; Bird et al., 2013), however the observations are not inconsistent with this picture (see, for example, Fumagalli et al., 2011; Bouché et al., 2013; Crighton et al., 2013).

There is thus a slew of evidence for a continuous interplay between multiple gas components at multiple temperatures, densities and metallicities, extending from the very centres of galaxies to beyond the virial radius. Indeed, tidal interactions between galaxies, as for example seen locally in the Magellanic stream (Mathewson et al., 1974; Brüns et al., 2005), indicate that these gaseous interactions extend to Mpc scales, and in a cosmological setting, galaxy overdensities are likely fuelled with gas falling, together with galaxies, along filaments on scales of tens of Mpc. Beyond this, a significant fraction of the IGM may also reside away from galaxy overdensities in galaxy voids (e.g. Penton et al., 2002; Tejos et al., 2012, 2014).

1.4 Rest frame observables

In this section, we describe the key rest-frame observables attributable to galaxies and the IGM.

1.4.1 Galaxy observables

Galaxies are characterised first and foremost by their apparent magnitudes, their colours, and their spectral features. Measuring a redshift also allows for a determination of their luminosities and rest-frame colours. These are the basic observables, which may be transformed into physical properties such as their star formation rates and stellar masses by making a number of prior assumptions on, for example, the initial stellar mass function, the star formation history, the density, temperature, and metallicity of the star-forming regions, the local ionization parameter, and the amount of dust in the galaxies. Some of these can be constrained from additional spectral information, but often these constraints are subject to further assumptions.

The spectra of galaxies consist of a relatively smooth stellar continuum, together with a collection of emission and absorption lines (which are not always present). The continuum is produced by the combination of many stellar spectra, which are approximately blackbody emitters spanning a range in temperature. The resulting continuum is therefore relatively flat. The emission lines are usually produced in the interstellar medium surrounding stars that are bright in the extreme ultraviolet (EUV). These are typically young and massive stars that are short-lived, and so the emission lines are almost always coincident with regions of active star formation. The EUV photons from these stars ionize the surrounding gas, and the resulting cascade of electrons through recombination is what gives rise to the emission lines. The most commonly observed rest-frame optical emission lines in galaxy spectra are the Hα and Hβ (Balmer series) lines, the [O II] $\lambda\lambda 3726, 3729$ and [O III] $\lambda\lambda 4959, 5007$ doublets, the [S II] $\lambda\lambda 6717, 6731$ doublet, and the [N II] $\lambda\lambda 6548, 6583$ doublet (e.g. Kennicutt, 1992). Galaxies with AGN often produce more highly ionized lines in addition to these, and they are some-

times much broader, owing to the larger velocity dispersions of the gaseous regions close to the super massive black holes (SMBHs) at their centres. In this case, it is the EUV and X-ray radiation from the AGN that ionizes the gas, rather than starlight, and this ionizing flux also produces extended transition regions from which low ionization lines are emitted. The absorption lines are usually produced in the atmospheres of typically old, low-mass stars, and in cold interstellar gas. A primary feature is the D4000 break, which is a break in the continuum at 4000 Å, and is caused by both the break in the Balmer series, and blanket absorption of radiation from metals in stellar atmospheres. In addition, the most common distinct absorption lines are the Ca II H and K lines, the CH molecule G-band line, the Mg λ5175 and Na λ5894 lines, and the Ca II $\lambda\lambda\lambda$8498, 8542, 8662 triplet. In the absence of Balmer emission lines, the corresponding Balmer absorption lines are also often visible.

In Figure 1.2 we show rest-frame composite spectra of a typical star-forming and non star-forming galaxy obtained from the Sloan Digital Sky Survey (SDSS; Abazajian et al., 2009). We label all the prominent emission and absorption lines. The differences between the spectra are striking. The star-forming galaxy spectrum shows an abundance of emission lines and a relatively flat continuum, with only a modest D4000 break, whilst the non star-forming galaxy spectrum is dominated by absorption lines, and shows a highly prominent D4000 break. Unsurprisingly then, the presence of optical emission lines (or lack thereof), and the prominence of the D4000 break, are often used as indicators of star formation activity. However, there are important caveats to bear in mind, such as the amount of dust in the galaxy, the underlying absorption, and the ionization parameter and metallicity in the ISM. All of these can affect the spectrum in different ways, and have to be taken into account.

The presence of dust in a galaxy will modify its spectrum by absorbing and scattering photons. This preferentially happens at shorter wavelengths, and so the net effect is often termed 'reddening'. Reddening by dust can be corrected for in a number of ways; one of the most common being the use of the Balmer decrement (e.g. Seaton, 1979; Cardelli et al., 1989; Calzetti, 2001), which is the measured

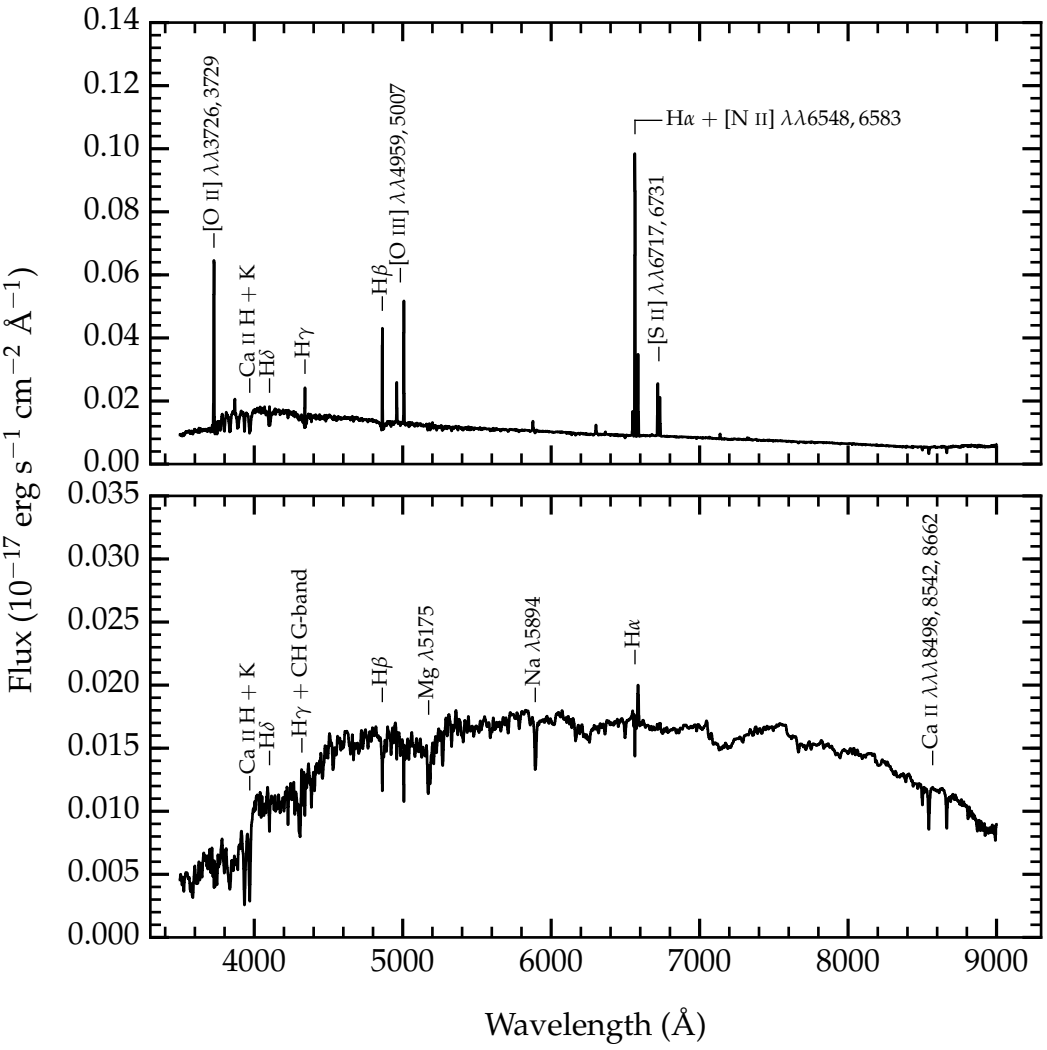

Figure 1.2: Composite rest-frame spectrum of a typical star-forming (top panel) and non star-forming (bottom panel) galaxy obtained from the Sloan Digital Sky Survey (SDSS; Abazajian et al., 2009), with some of the most prominent emission and absorption lines labelled.

ratio of Hα and Hβ line strengths. By comparing this ratio to that expected in theory (for a assumed density and temperature; Osterbrock, 1989), and combining this with a model for the wavelength-dependence of the attenuation (e.g. Calzetti, 2001; Fischera et al., 2005), a 'dust-free' spectrum can be recovered.

Underlying stellar absorption beneath emission lines leads to measured line fluxes that are lower than the intrinsic ones. This particularly affects the Balmer lines, and can be corrected for by considering the absorption at weaker lines in the series (where the absorption component may dominate, e.g. at Hδ; Hopkins et al.,

2003), or by fitting templates generated from stellar population synthesis models (e.g. Bruzual & Charlot, 2003; Maraston & Strömbäck, 2011).

Uncertainty due to the ionization parameter and metallicity in the ISM is typically minimised by considering emission lines that are less sensitive to these parameters, although they can be estimated to some degree by considering ratios between emission lines in combination with photoionization and spectral energy distribution (SED) modelling (e.g. Kewley et al., 2001). Calculations such as these are also an important diagnostic on the contribution from AGN activity, which can dominate substantially over that from star formation. Spectra that are dominated by an AGN component are usually discarded in the construction of star-forming samples, since it is often very difficult (or impossible) to subtract away this contribution.

1.4.2 IGM observables

Due to the extremely low density of the IGM, its observation is currently limited and difficult. In most circumstances, the only way to characterise it is through the measurement of intervening absorption lines or flux decrements in QSO spectra,[5] as was first proposed and implemented in the mid 1960s (Gunn & Peterson, 1965; Scheuer, 1965; Lynds & Stockton, 1966; Burbidge et al., 1966; Stockton & Lynds, 1966; Kinman et al., 1966). This limits the characterisation of the IGM to being one-dimensional, but for spectra with reasonable signal-to-noise ratio (SNR) ($\gtrsim 10$), allows for the detection of gas with very low column densities ($\log(N/\mathrm{cm}^{-2}) \gtrsim 13$) in neutral hydrogen. Similarly low column densities may also be probed in some metal transitions (see Chapter 2, Section 2.6 for an exhaustive overview). These column densities are many orders of magnitude below that currently accessible with emission-line studies, which are limited to the densest gas in and around the immediate environments of galaxies and sites of star formation.

We show an illustration of the QSO absorption line technique in Figure 1.3. For

[5]There are also recent efforts to do the same with high redshift galaxies that are bright in the rest-frame ultraviolet (e.g. Lee et al., 2014), although these galaxies are far fainter than QSOs, which makes this approach currently very challenging.

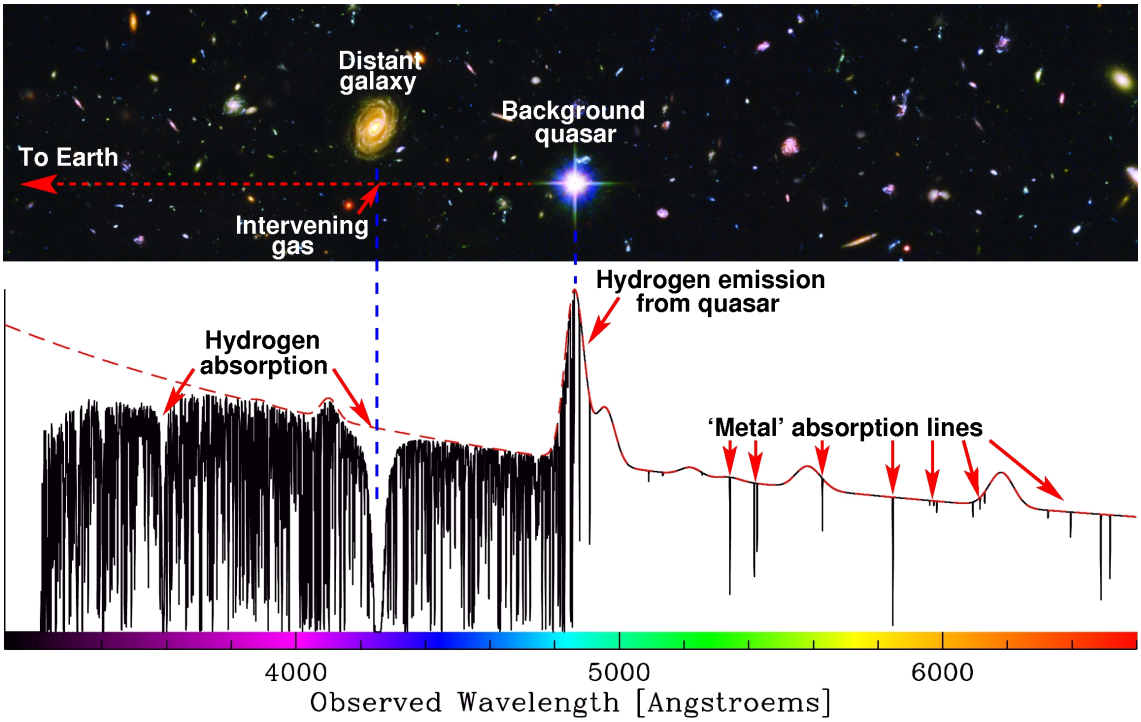

Figure 1.3: An illustration of the QSO absorption line technique. Neutral hydrogen along the line-of-sight to a background QSO creates a 'forest' of absorption lines at wavelengths shorter than the observed wavelength of the QSO Lyα emission line, known as the Lyα forest. Absorption lines at wavelengths longer than the Lyα emission line are attributable to intergalactic metals. © Michael Murphy.

$z \gtrsim 3$, the Lyα emission line of atomic hydrogen is redshifted into the optical window, and is accessible from ground-based telescopes and instrumentation. Being by far the most abundant element in the Universe, most of the absorption lines in QSO spectra at wavelengths less than that of the redshifted Lyα line are expected to come from the hydrogen Lyα transition. This is indeed what is observed, with there being an apparent 'forest' of absorption lines at these wavelengths, commonly known as the Lyα *forest*. Other absorption lines at wavelengths greater than the redshifted Lyα line are also present, attributable to intergalactic metals, although these lines are far less common.

Fundamentally, the absorption of photons from a distant source is due to the presence of atoms and electrons in the IGM that scatter the light. If we consider a beam of light with intensity I, falling on a plane-parallel slab of material with thickness s and number density n, then if some of the light is scattered by the

material away from the beam, we can write

$$\frac{\mathrm{d}I}{I} = -\sigma n \, \mathrm{d}s, \tag{1.4.29}$$

where σ is some constant that depends on the material, having dimensions of surface area. For a constant number density, we can integrate this to find

$$I = I_0 \mathrm{e}^{-\sigma n s} = I_0 \mathrm{e}^{-\tau}, \tag{1.4.30}$$

where

$$\tau = \sigma n s \tag{1.4.31}$$

is the *optical depth*, and I_0 is the unattenuated intensity. The constant σ then describes the strength of the interaction between light and the particles that make up the slab.

A simple interaction between matter and radiation is the scattering of photons by free electrons - *Thomson scattering*. In this process, the electric field of the incoming electromagnetic radiation exerts an oscillating force on the electron, causing it to oscillate and radiate energy. In the classical limit, the fraction of scattered light is independent of frequency, and the cross-section is the *Thomson cross-section*:

$$\sigma_\mathrm{T} = \frac{8\pi}{3}\left(\frac{e^2}{m_e^2 c^2}\right)^2, \tag{1.4.32}$$

where e is the electron charge, m_e is the electron mass, and c the speed of light. This can be substituted for the σ in the equations above to give the optical depth for Thomson scattering.

Now let's consider the scattering induced by a photon interacting with an electron in a harmonic potential well (a harmonic oscillator). The equation of motion is

$$m_e \ddot{x} = -m_e \omega_0^2 x + m_e \zeta \dot{x} + e E_0 \sin(\omega t), \tag{1.4.33}$$

where ω_0 is the frequency of the harmonic oscillator, ζ represents a damping term, and E_0 is the amplitude of the electric field associated with the external radiation field, driven with frequency ω. In this case, the cross-section for scattering depends on the frequency of the incoming radiation, $\nu = 2\pi\omega$, and the harmonic oscillator will radiate most efficiently when in resonance with the incoming radiation, i.e. when $\omega \approx \omega_0$:

$$\begin{aligned}\sigma(\nu) &= \frac{\pi e^2}{m_e c}\frac{4\gamma}{16\pi^2(\nu-\nu_0)^2+\gamma^2} \\ &= c\sqrt{\frac{3\pi\sigma_{\mathrm{T}}}{8}}\frac{4\gamma}{16\pi^2(\nu-\nu_0)^2+\gamma^2} \\ &\equiv \sigma_0\phi(\nu),\end{aligned} \tag{1.4.34}$$

where $\gamma = \omega_0^2\zeta$. We note that $\int_0^\infty \mathrm{d}\nu\ \phi(\nu) = 1$, so σ_0 is the net cross-section of the transition, and ϕ describes the line shape:

$$\phi(\nu) = \frac{4\gamma}{16\pi^2(\nu-\nu_0)^2+\gamma^2}, \tag{1.4.35}$$

which is called the *Lorentz profile*.

In reality, we're concerned with interactions of photons with bound states of atoms, whereby an atom in a lower energy state, l, can absorb a photon of energy $E_{lu} = E_u - E_l$, and attain an upper energy state, u. The cross-section is frequency dependent, but strongly peaked at the frequency $\nu_0 = \nu_{lu} = E_{lu}/h$ (where h is Planck's constant), and is accurately given by

$$\sigma(\nu) = \frac{\pi e^2}{m_e c} f_{lu}\frac{4\gamma_{lu}}{16\pi^2(\nu-\nu_{lu})^2+\gamma_{lu}^2}. \tag{1.4.36}$$

This is reduced from the cross section for a damped, driven harmonic oscillator in equation (1.4.34) by a factor f_{lu}, referred to as the f-value for the transition, otherwise known as the *oscillator strength*, and corresponds to the integral over the product of wave-functions of the l and u states. The natural line width, γ_{lu}, ap-

pears as a result of the Heisenberg uncertainty principle.[6] Because the line shape is so similar to that of a damped, driven harmonic oscillator, the extended wings of the Lorentz profile, which result from the damping term, are often called *damping wings*.

Now let's suppose the IGM is homogeneous, with proper density $n(z)$. We let $\rho_{\rm crit} = 3H_0^2/(8\pi G)$ be the present day critical density, and take $y \approx 0.24$ to be the helium mass fraction (from Big Bang nucleosynthesis). From equation (1.4.31), the mean number density of hydrogen atoms at a redshift z is then

$$n_{\rm H} = n_0 a^{-3}, \tag{1.4.37}$$

where

$$n_0 \equiv 1.7 \times 10^{-7} \frac{\Omega_{\rm b} h^2}{0.02}\ {\rm cm}^{-3}, \tag{1.4.38}$$

and $\Omega_{\rm b}$ is the baryon contribution to the dimensionless density parameter. For $z \geq 1$, we can assume $H(a) \approx H_0 \Omega_{\rm m}^{1/2} a^{-3/2}$, with $\Omega_{\rm m} \approx 0.3$. We'll also assume that the Lyα cross-section is very peaked, so that $\sigma(\nu) \approx \sigma_0 \delta(\nu - \nu_0)$, where δ is the Dirac delta function. The optical depth at a frequency ν is then

$$\begin{aligned}
\tau(\nu) &= \int_a^1 \sigma(\nu/a) n(a) \frac{c\ {\rm d}a}{aH(a)} \\
&= \frac{\sigma_0 n_0 c}{H_0 \Omega_{\rm m}^{1/2}} \int_a^1 {\rm d}a\ \delta(\nu/a - \nu_0) a^{-5/2} \\
&= \frac{\sigma_0 n_0 c}{H_0 \Omega_{\rm m}^{1/2}} \frac{a^{-3/2}}{\nu_0} \\
&= 13\,000 h^{-1} \frac{\Omega_{\rm b} h^2}{0.02} (1+z)^{3/2}.
\end{aligned} \tag{1.4.39}$$

Inserting measured values for the above parameters implies the optical depth in Lyα at $z \sim 3$, for example, should be around 10^5. If this were the case, we shouldn't expect to see any transmitted flux below the Lyα emission line at all. The only

[6] The Heisenberg uncertainty principle, $\Delta E \Delta t \geq \hbar$, implies that an energy level u has a width $\Delta E \approx \hbar/\tau_u$, where τ_u is the level lifetime.

reason we do see the Lyα forest is because the Universe is highly ionized, with neutral hydrogen contributing only one part in every $\sim 10^5$ to the total or less!

To obtain a quantitative physical description of the intergalactic matter that gives rise to QSO absorption lines, we seek a theoretical description for the absorption line profiles. In reality, this is not just the Lorentz profile in equation (1.4.35), because we have thus far neglected thermal motions in the gas. According to the Boltzmann distribution, for a gas in thermodynamic equilibrium, the fraction of particles with mass m and velocity in the interval $[v, v + \mathrm{d}v]$ is

$$N(v) = \frac{1}{\sqrt{2\pi kT/m}} \mathrm{e}^{-v^2/(2kT/m)}. \tag{1.4.40}$$

The fraction of photons absorbed at a Doppler shifted frequency ν is just proportional to the fraction of particles with velocity $v = c(\nu/\nu_{lu} - 1)$. Therefore the line shape from thermal motions alone is

$$\begin{aligned} \phi(\nu) &= N(v)\frac{\mathrm{d}v}{\mathrm{d}\nu} \\ &= \frac{c}{\nu_{lu}} \frac{1}{\sqrt{2\pi kT/m}} \exp\left(-\frac{(\nu - \nu_{lu})^2}{2\nu_{lu}^2 kT/mc^2}\right), \end{aligned} \tag{1.4.41}$$

which we rewrite as

$$\phi(\nu) = \frac{1}{\sqrt{\pi \Delta\nu_{\mathrm{D}}^2}} \exp\left(-\frac{(\nu - \nu_{lu})^2}{\Delta\nu_{\mathrm{D}}^2}\right), \tag{1.4.42}$$

where

$$\Delta\nu_{\mathrm{D}} \equiv \frac{\nu_{lu}}{c}\sqrt{\frac{2kT}{m}}. \tag{1.4.43}$$

The net profile describing an absorption line, the *Voigt profile*, is then a convolution between the Lorentz profile in equation (1.4.35), and the thermal Gaussian profile (Maxwell-Boltzmann profile) in equation (1.4.42):

$$\phi(\nu) = \int_{-\infty}^{\infty} \mathrm{d}v \, \frac{1}{\sqrt{\pi} b} \exp\left(-\frac{v^2}{b^2}\right) \frac{4\gamma_{lu}}{16\pi^2(\nu - \nu_{lu}(1 - v/c))^2 + \gamma_{lu}^2}, \tag{1.4.44}$$

where

$$b \equiv \sqrt{\frac{2kT}{m}} = \frac{c\Delta\nu_{\rm D}}{\nu_0} \tag{1.4.45}$$

is the *Doppler broadening parameter* (km s^{-1}). In general, absorption lines may also be broadened by turbulent motions, so when referring to the Doppler broadening parameter, we typically mean

$$b^2 = b_{\rm turb}^2 + \frac{2kT}{m}, \tag{1.4.46}$$

where $b_{\rm turb}$ parametrises the turbulent contribution to the line broadening (assumed Gaussian).

Finally, the optical depth in an absorption line may be written

$$\tau(\nu) = \frac{\pi e^2}{m_e c} f_{lu} N_l \phi(\nu), \tag{1.4.47}$$

where

$$N_l \equiv \int {\rm d}s \; n_l \tag{1.4.48}$$

is the *column density* (cm^{-2}), and n_l is the number density of atoms in energy state l. The dependence on the observed frequency, ν, can be re-cast into a dependence on velocity, v, via

$$\nu = \frac{\nu_{lu}}{1+z}\left(1 - \frac{v}{c}\right). \tag{1.4.49}$$

An absorption line may therefore be parametrised by three quantities: a redshift, z, a column density,[7] N, and a Doppler broadening parameter, b. Fitting a Voigt profile to an observed absorption line therefore allows one to estimate these quantities. In addition, for a Gaussian velocity distribution (equation (1.4.42)), taking

[7]For the majority of this thesis, we shall drop the subscript l.

the limit $\nu \to \nu_{lu}$, we have that the *optical depth at line centre* is

$$\tau_0 = \sqrt{\pi}\frac{e^2}{m_e c}\frac{N_l f_{lu}\lambda_{lu}}{b}, \tag{1.4.50}$$

where $\lambda_{lu} = c/\nu_{lu}$ is the rest-frame wavelength for the atomic transition of interest.

When taking a spectrum of a distant source, what is actually being measured is the flux density, F_ν, as a function of ν (equation (1.2.25)). We observe the source using an aperture of solid angle $\Delta\Omega$, and assume the properties of the foreground gas are essentially uniform over $\Delta\Omega$. Integrating equation (1.4.30) over the (small) solid angle $\Delta\Omega$ we obtain

$$F_\nu = F_\nu(0)\mathrm{e}^{-\tau}, \tag{1.4.51}$$

where $F_\tau(0)$ is the flux density from the source in the absence of absorption. To measure $\tau(\nu)$, one therefore requires an estimate for the SED of the source over the range of interest (see Chapter 2, Section 2.5 for details). Using this estimate, one can also calculate the *dimensionless equivalent width* of an absorption line:

$$W \equiv \int \frac{\mathrm{d}\nu}{\nu_{lu}}\left(1 - \frac{F_\nu}{F_\nu(0)}\right) = \int \frac{\mathrm{d}\nu}{\nu_{lu}}(1 - \mathrm{e}^{-\tau}), \tag{1.4.52}$$

which is also sometimes quoted as the wavelength equivalent width:

$$W_\lambda \equiv \int \mathrm{d}\lambda\ (1 - \mathrm{e}^{-\tau}) \approx \lambda_{lu} W. \tag{1.4.53}$$

In this thesis, we shall mainly make use of the latter.

To conclude this section, it is instructive to examine various limits in τ. In the optically thin regime, $\tau \ll 1$, and expanding $(1 - \mathrm{e}^{-\tau})$, we have

$$\begin{aligned} W &\approx \sqrt{\pi}\frac{b}{c}\tau_0\left(1 - \frac{\tau_0}{2\sqrt{2}} + \ldots\right) \\ &\approx \sqrt{\pi}\frac{b}{c}\frac{\tau_0}{1 + \tau_0/(2\sqrt{2})} \\ &= \frac{\pi e^2}{m_e c^2} N_l f_{lu}\lambda_{lu}\frac{1}{1 + \tau_0/(2\sqrt{2})}. \end{aligned} \tag{1.4.54}$$

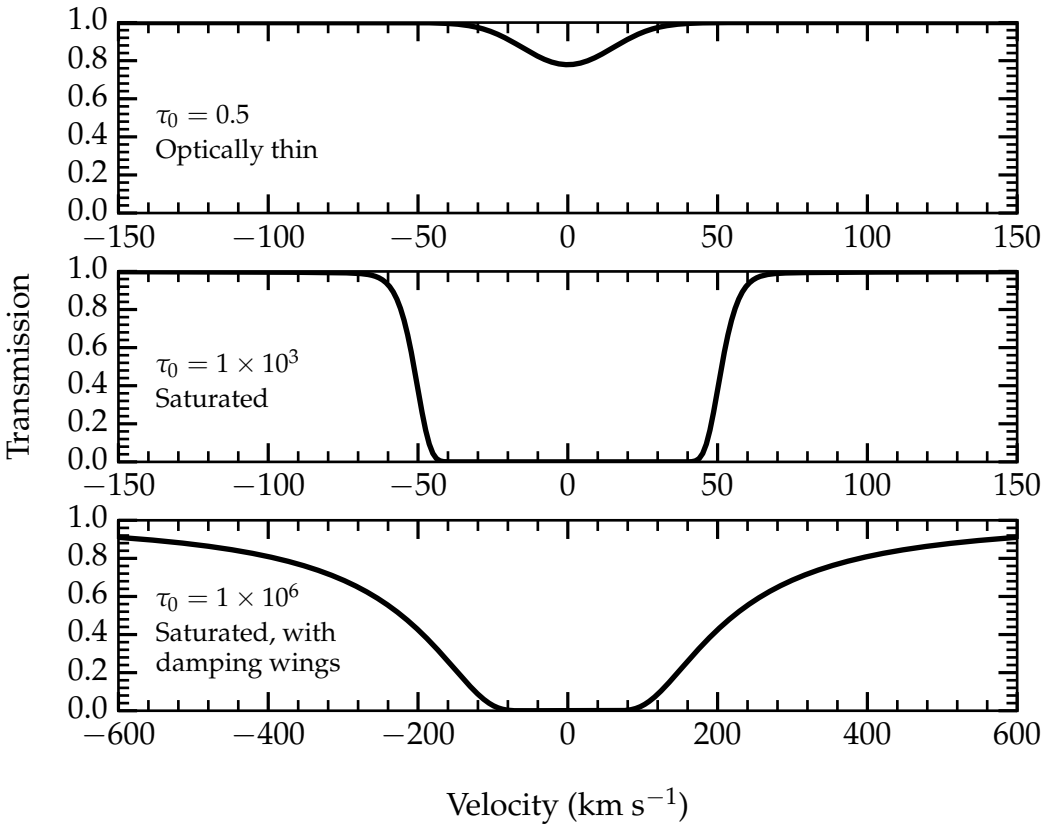

Figure 1.4: Absorption line profiles in three regimes, using the H I Lyα transition as an example, with $b = 20$ km s^{-1} and $N(\text{H I}) = 6.6 \times 10^{12}$ cm^{-2}, $N(\text{H I}) = 1.3 \times 10^{16}$ cm^{-2}, and $N(\text{H I}) = 1.3 \times 10^{19}$ cm^{-2} in the upper, middle, and lower panels respectively. Note the range on the x-axis in the bottom panel.

This is exact for $\tau_0 \to 0$, and we find that

$$N_l = 1.130 \times 10^{12} \text{ cm}^{-1} \frac{W}{f_{lu}\lambda_{lu}} \qquad (\tau_0 \ll 1). \tag{1.4.55}$$

Therefore, even if we do not resolve an absorption line that is optically thin, measurement of W allows for determination of N_l. An example of an absorption line in this regime is shown in the top panel of Figure 1.4. Away from the optically thin regime, equation (1.4.55) gives a lower limit on N_l.

As τ increases, and the core of the absorption line starts to saturate, the quantity $(1 - e^{-\tau})$ becomes increasingly 'box-shaped', as is depicted in the middle panel of Figure 1.4. If we treat the optical depth as a delta function with only Doppler

broadening, i.e.

$$\tau(v) = \tau_0 e^{-(v/b)^2}, \tag{1.4.56}$$

then we can approximate W in this regime by the fractional full width at half maximum (FWHM):

$$W \approx \frac{(\Delta\nu)_{\rm FWHM}}{\nu_{lu}} = \frac{(\Delta v)_{\rm FWHM}}{c} \approx \frac{2b}{c}\sqrt{\ln(\tau_0/\ln 2)}. \tag{1.4.57}$$

In this case, W is *very* insensitive to τ_0 (and therefore N_l), which makes determining N_l extremely challenging.

As τ increases yet further, the core of the absorption line is completely saturated, but the damping wings in the profile start to become apparent (see the bottom panel of Figure 1.4). In the limiting case, we can entirely neglect the Doppler broadening of the line, and assume that the wings follow a pure Lorentz profile (equation (1.4.35)):

$$\tau(\nu) \approx \frac{\pi e^2}{m_e c} N_l f_{lu} \frac{4\gamma_{lu}}{16\pi^2(\nu-\nu_{lu})^2 + \gamma_{lu}^2} \qquad (|\nu - \nu_{lu}| \gg \nu_{lu} b/c), \tag{1.4.58}$$

and

$$W = \sqrt{\frac{b}{c}\frac{\tau_0}{\sqrt{\pi}}\frac{\gamma_{lu}\lambda_{lu}}{c}}. \tag{1.4.59}$$

We are back to a linear dependence of N_l on W in this regime, and determining the column density is once again straightforward.

In Figure 1.5, we plot W as a function of the product $N_l f_{lu} \lambda_{lu}$ for different b-values in the range $10 - 100$ km s^{-1}, where we have taken $\gamma_{lu}\lambda_{lu}$ as appropriate for the Lyα transition of atomic hydrogen. This function is known as the *curve of growth*. The region where the curves for different b-values separate is described by equation (1.4.57), and is known as the *flat portion* of the curve of growth. Here, the difficulty in measuring N_l is very much apparent. Numerous combinations of N_l and b can give equivalent widths that are indistinguishable. The situation

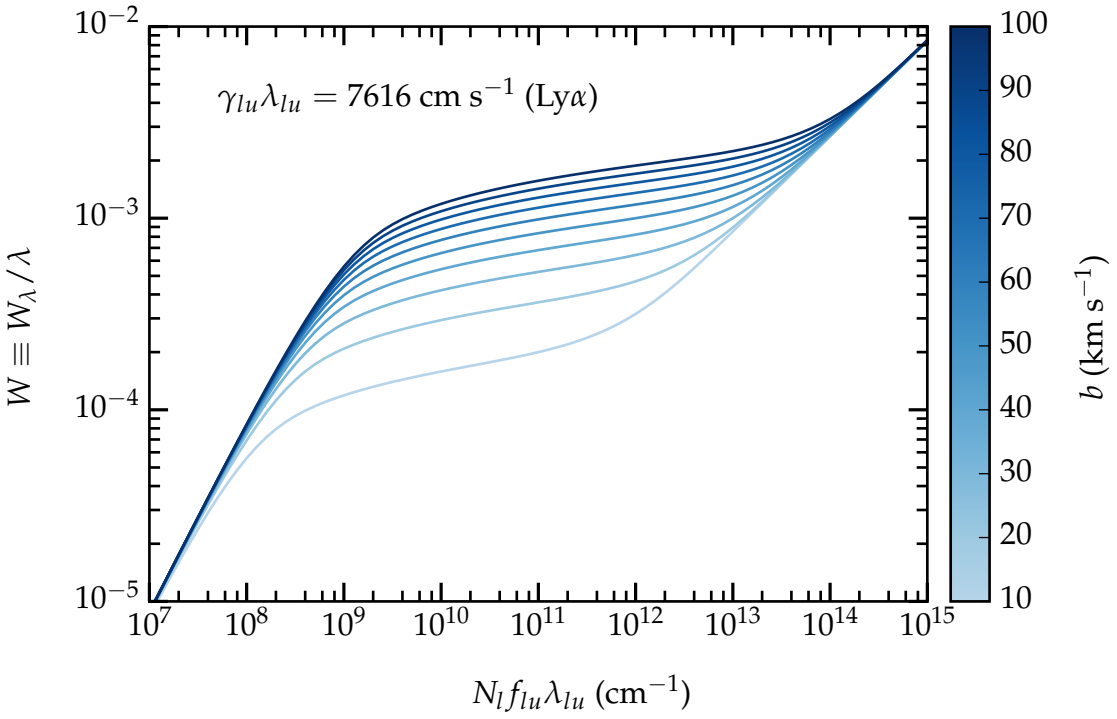

Figure 1.5: The curve of growth as a function of Doppler broadening parameter in the range 10 – 100 km s^{-1} for a damping constant $\gamma_{lu}\lambda_{lu} = 7616$ cm s^{-1} (appropriate for H I Lyα).

is usually helped somewhat when multiple transitions of a given ion are available. These will have different f_{lu}, λ_{lu}, and γ_{lu}, and therefore different curves of growth, thus providing a better constrained solution for N_l and b. The region on the plot having smaller values of $N_l f_{lu} \lambda_{lu}$ corresponds to equation (1.4.54), and is referred to as the *linear portion* of the curve of growth. The region with higher values of $N_l f_{lu} \lambda_{lu}$ corresponds to equation (1.4.59), and is referred to as the *damped portion* of the curve of growth.

Draine (2011) provides a useful approximation formula for $W(\tau_0, b, \gamma_{lu}, \lambda_{lu})$, which is continuous and accurate to within a few percent:

$$W \approx \begin{cases} \sqrt{\pi}\frac{b}{c}\frac{\tau_0}{1+\tau_0/(2\sqrt{2})} & (\tau_0 < 1.25393), \\ \left[\left(\frac{2b}{c}\right)^2 \ln\left(\frac{\tau_0}{\ln 2}\right) + \frac{b}{c}\frac{\gamma_{lu}\lambda_{lu}}{c}\frac{(\tau_0 - 1.25393)}{\sqrt{\pi}}\right]^{1/2} & (\tau_0 > 1.25393). \end{cases} \tag{1.4.60}$$

1.5 Numerical simulations

Galaxies and their environments are extremely complicated dynamical systems, and it is usually very difficult to obtain analytical solutions to the equations governing them unless we make a great number of simplifying assumptions. Therefore, much use has been made of numerical simulations to guide the interpretation of observational data. This has been made possible by the rapid increase in the availability and capability of computer hardware and software in recent decades. The simulations are now large enough to capture the non-linear growth of cosmic structures, and to test the validity of the ΛCDM paradigm for galaxy formation.

In a numerical simulation, the distribution of matter is usually represented either as a set of particles (a *Lagrangian* approach), or on a grid (an *Eulerian* approach). The motion of each mass element is then traced numerically by taking into account its interactions with all other mass elements at each time step.[8] If we could take into account the motions of all elementary particles, with all the relevant forces, simulations of this kind would give exact results. However, this is far from realistic bearing in mind the computing power available today, with a single galaxy such as our Milky Way containing $\sim 10^{68}$ protons! In practice, pseudo-particles or mass elements are used to represent the distribution of matter, each having a mass many orders of magnitude larger than that of an atom, or in simulations over cosmologically significant volumes, many order of magnitude larger than that of a star. From a pessimistic viewpoint, this may represent a serious limitation on the reliability of these simulations.

Initially, numerical simulations were concerned mostly with addressing the gravitational forces at work in shaping the large-scale matter distribution (see Davis et al., 1985, for one of the earliest examples). Gravity is the only force that acts non-negligibly on the hypothesised (cold) dark matter that makes up nearly 85% of the matter in the Universe, so in simulating just the evolution of dark matter, one can neglect all other forces. These so-called 'N-body' simulations (Hock-

[8] In general, the most computationally efficient codes only calculate interactions between individual particle or grid cells pairs at small distances, with interactions between groups of particles or grid cells providing an accurate approximation at large distances.

ney & Eastwood, 1981) are able to address many of the key questions in the field of cosmology, such as those regarding the formation of large scale structures, and the abundance of dark matter haloes. However, to simultaneously simulate the evolution of baryonic structures in the Universe, one also needs to solve the equations of hydrodynamics.[9] Those simulations that solve both the equations of gravity and hydrodynamics as a function of time, starting from an initial mass distribution set by cosmological initial conditions constrained from observations of the CMB, are referred to as *cosmological hydrodynamical simulations*.

In principle, cosmological hydrodynamical simulations can follow the evolution of both baryons and dark matter without relying on simplified approximations for all the important processes. However, in practice, a number of approximations *are* still required due to the limited numerical resolution. These approximations are typically a set of 'prescriptions' or 'recipes' that are conditionally executed at each time step to capture physics that is not resolved by the interactions of mass elements alone. Since these processes occur below the resolution of the simulation 'grid', the implementation of these processes is commonly referred to as the *'subgrid' physics*. As an example, the detailed star-formation processes within cold, dense molecular clouds are not directly captured by the simulations. Therefore, the star-formation rate within a fluid element is usually modelled according to the empirical Kennicutt-Schmidt law (Kennicutt, 1998b), based on the local cold gas density (or pressure). Even the cold molecular gas regions themselves may not be resolved, requiring there to be a critical density threshold above which cold molecular gas is expected to form.

In the simulations, galaxies are usually identified as massive and dense collections of stars, whose luminosities and colours are obtained from their star formation histories (which are known), together with the use of stellar population synthesis models (e.g. Bruzual & Charlot, 2003; Maraston & Strömbäck, 2011), which predict the resulting galaxy SEDs. A common problem has been the failure to adequately reproduce the galaxy stellar mass functions inferred from observations

[9]The equations governing fluid flows.

(e.g. Li & White, 2009; Baldry et al., 2012). This occurs predominantly when too much gas is allowed to cool and form stars, and has been termed the 'overcooling problem' (e.g. Katz et al., 1996; Balogh et al., 2001; Kereš et al., 2009a; Schaye et al., 2010). To overcome this, a set of processes collectively termed 'feedback' are required to efficiently couple the energy, mass, and metals released by supernovae and AGN to the surrounding gas, which acts to regulate gas cooling and suppress the rate of star formation.[10] In particular, feedback from supernovae is required to suppress the formation of low-mass galaxies, and feedback from AGN is required to suppress the formation of high-mass galaxies. Unfortunately, much like star formation, feedback from supernovae and AGN cannot be treated from first principles in the simulations, and so they are typically implemented as a set of energy/momentum sources, whose efficiencies are simply calibrated to match statistical properties of the galaxy population, such as the galaxy stellar mass function, and the relationship between stellar mass and SMBH mass (e.g. McConnell & Ma, 2013).

Due to their reliance on uncertain subgrid physics, the predictive power of cosmological hydrodynamical simulations is currently limited, although they do provide a promising avenue for the study of baryonic structure formation and evolution. Currently however, comparisons to observational data, and the interpretations drawn from the simulations, are often treated with caution. It is therefore of great interest to better constrain the subgrid physics that must inevitably be included at present.

1.6 Motivation and structure of the thesis

In this thesis, we aim to advance our quantitative understanding on the relationship between the IGM and galaxies. The broad expectation from theory is that galaxies form stars via the accretion of cold gas along filamentary streams and/or via condensations from their hot coronae. Juxtaposed against this, feedback from

[10]Note that there is a subtly here, as the injection of metals actually acts to increase the rate at which gas radiatively cools.

supernovae and AGN redistributes metal-enriched gas back into the IGM. We shall test this picture using QSO absorption line methods in combination with spectroscopic surveys of galaxies to observationally constrain the dynamics and distribution of metal-enriched gas in and around galaxies on scales of a few kpc, to tens of Mpc. By computing correlation statistics, we will statistically characterise the gas-galaxy connection over a large range of environments, from galaxy halo environments, to regions surrounding galaxy groups, clusters and filaments, and out to regions almost devoid of galaxies. These measurements will also provide a set of observational constraints on the subgrid feedback prescriptions implemented in cosmological hydrodynamical simulations that are potentially orthogonal to those used to calibrate them in the first place. Comparing these results with the simulations should therefore glean crucial insights on the physics driving the evolution of galaxies and baryonic structure in the Universe.

The structure of this thesis is as follows: In Chapter 2, we present the QSO sample and describe the methods used to construct a sample of IGM absorbers. In Chapter 3, we present a detailed analysis of the intrinsic absorption associated to one of these QSOs, which provides powerful constraints on the nature of feedback. In Chapter 4, we present the spectroscopic survey of galaxies constructed for this work. In Chapter 5, we describe the creation of a set of comparison data drawn from the Evolution and Assembly of GaLaxies and their Environments (EAGLE) cosmological hydrodynamical simulation. In Chapter 6, we present the results of a cross-correlation study between galaxies and O VI absorbers, which we use to characterise the distribution and dynamics of metal-enriched gas around galaxies. These results are then compared to an identical calculation performed using the data set extracted from the EAGLE simulation. Finally, in Chapter 7, we discuss the implications of these results and draw conclusions.

Chapter 2

Detecting the IGM in absorption

2.1 Overview

In this thesis, we make use of quasar (QSO) absorption-line data obtained from the Hubble Space Telescope (*HST*)/Cosmic Origins Spectrograph (COS) and Faint Object Spectrograph (FOS) to study the low redshift ($z < 1$) intergalactic medium (IGM). This chapter presents a description of the data reduction and analysis leading to the creation of an absorption line list, which forms the basis for much of the work that follows. We begin in Section 2.2 with an overview of COS, the primary instrument employed for this work. The full sample of QSOs used in subsequent sections of this thesis is summarised in Section 2.3, including many that have been obtained, reduced, and analysed by other authors. In Section 2.4 we describe the data reduction procedures. Analysis of our own fully reduced QSO set is then described in Sections 2.5 to 2.7. These cover psuedo continuum fitting, absorption line identification, and absorption line fitting. We conclude with a description of our absorber sample in Section 2.8.

2.2 The Cosmic Origins Spectrograph

The Cosmic Origins Spectrograph is a UV spectrograph that was installed on the Hubble Space Telescope in May 2009. It was designed, for the most part, to dramatically increase the number of QSOs available for spectroscopic observation at UV wavelengths, leading to a much denser sampling of the IGM in absorption at low redshifts ($z < 1$; Green et al., 2012). The specific goal was to be 10 times more efficient at gaining the same signal-to-noise ratio (SNR) for a given object

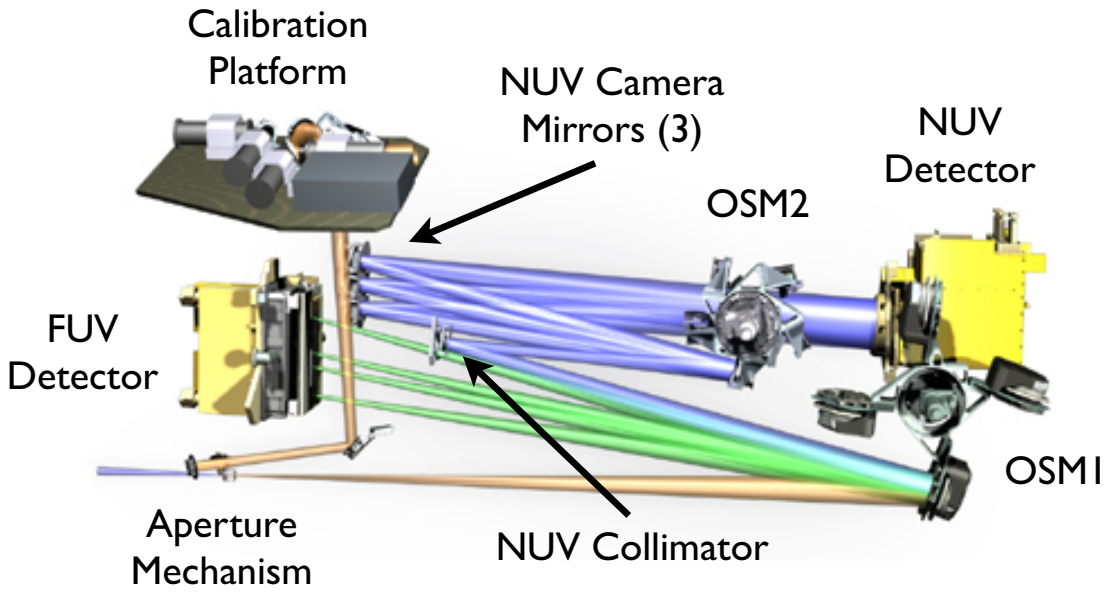

Figure 2.1: The layout of COS. The optical path is drawn to scale with all items in proportion and in their correct locations relative to one another. Sourced from the COS Instrument Handbook (Debes et al., 2015).

(using ~ 1/10th the observing time) compared to the previous generation ultraviolet (UV) spectrograph, the Space Telescope Imaging Spectrograph (STIS). This leads to 1000 times as many QSOs available for observation in a typical integration time.

COS is a slit-less spectrograph designed for point-source spectroscopy, and uses two circular 2.5 arcsec diameter science apertures: the primary science aperture (PSA), which is open, and the bright object aperture (BOA), which contains a neutral density filter to reduce the flux from bright objects. It has two channels: a far ultraviolet (FUV) channel covering wavelengths of 900–2150 Å, and a near ultraviolet (NUV) channel covering wavelengths 1650–3200 Å, offering both moderate resolution ($R = \lambda/\Delta\lambda \simeq 18\,000$) and low resolution ($R \simeq 3000$) spectroscopy. The NUV channel also offers limited imaging capability.

The purpose of the small science aperture was to minimise the light contribution from the sky (i.e. geocoronal emission) and any potential confusion with nearby point sources. It is set to lie at the minimum waist point in the *HST* optical tube assembly (OTA) point-spread function (PSF), which is the point at which 100% of the transmitted light is included in the minimum possible diameter. Fo-

cusing COS on the sky results in the science aperture itself being out of focus, and so the transmission curve from the sky as a function of off-axis angle does not have a sharp cut-off (see figure 2 in Green et al., 2012). In reality, the science aperture passes slightly less than 100% of the original beam due to uncorrectable aberrations attributable to the *HST* optical tube assembly, as well as scattered light due to dust and surface imperfections. The layout of COS is illustrated in Figure 2.1.

2.2.1 FUV design

The FUV channel utilises just one grating and no other optical elements. This choice was motivated by the low reflectivity of surfaces at FUV wavelengths (at best $\sim$ 80%). Since COS is a point-source spectrograph, limiting the field of view to 2.5″ removes the need for multiple optics to correct for the spherical aberration in the *HST* OTA and the path-length errors introduced by the primary mirror, such as those in STIS (Woodgate et al., 1998) and the Advanced Camera for Surveys (ACS) (Ford et al., 1998). All these corrections are instead performed by a single grating, in addition to its primary function of dispersing the light entering the spectrograph. A slit-less design for COS is therefore key to ensuring that transmission is maximal.[1] The layout of the FUV optical path is shown in Figure 2.2.

COS focuses on the sky rather than on a slit, so that the spectral width of any monochromatic point source is only as large as the corrected *HST* line-spread function (LSF) (Ghavamian et al., 2009; Kriss, 2011), plus any aberrations in the spectrograph. These aberrations are minimised in the design of COS by using a holographic grating, which also delivers the desired spectral resolution. A holographic grating is one formed by exposing a polished substrate coated with a light-sensitive material (photoresist) to an interference pattern produced from two laser beams. The resulting pattern is one of straight lines with a sinusoidal cross-section. Traditionally, holographic gratings have the advantage of producing less scattered light than gratings that are mechanically cut, but they lose out on the

[1]Without correcting optics in front of the exit pupil, a point source does not form a well-defined image with *HST*, so light is lost in the presence of a narrow slit.

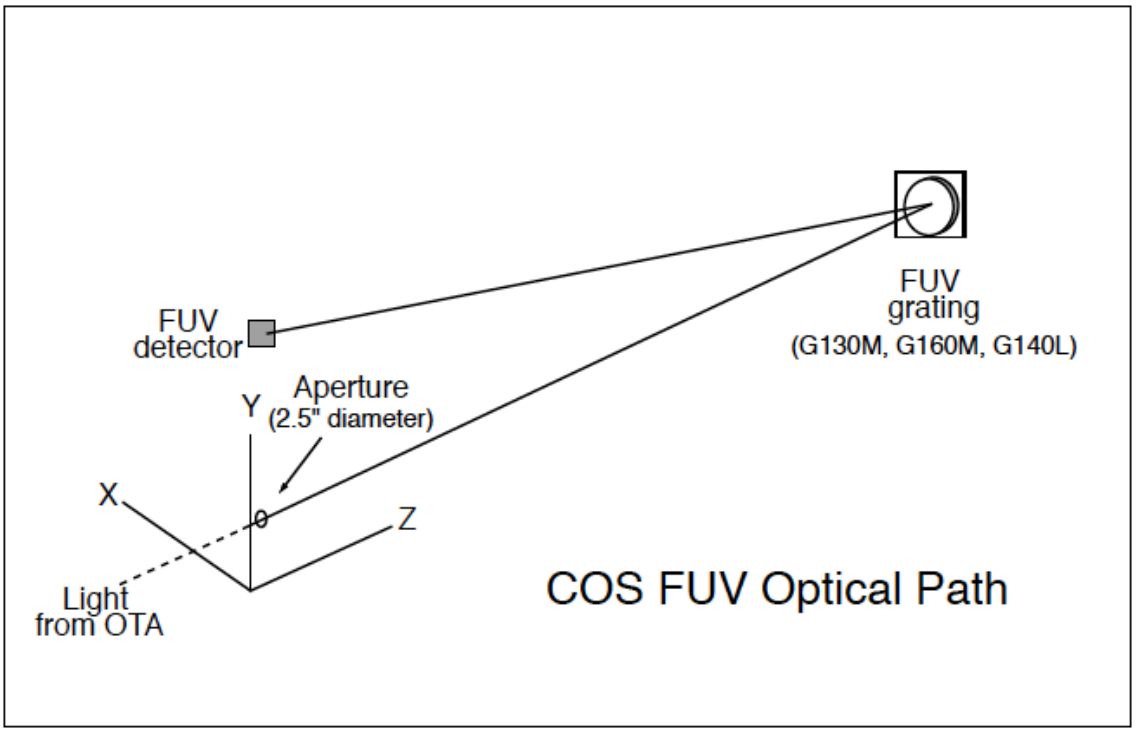

Figure 2.2: The COS FUV optical path, sourced from the COS Instrument Handbook (Debes et al., 2015).

high groove efficiency[2] of the latter. However, the COS FUV grating design has taken advantage of techniques that now exist for creating a blazed groove profile on a holographic grating, produced by exposing the grating to an ion-etch beam at different angles for different lengths of time. The result is a holographic grating with a triangular blazed profile (see figure 3 in Green et al., 2012). The FUV holographic gratings are coated on a roughly spherical substrate with polynomial deformations in the fourth and sixth orders. This corrects for spherical aberrations and the deviations in the *HST* primary mirror from its intended hyperbolic shape.

The FUV channel on COS utilises a large-format detector consisting of two 85×10 mm microchannel plate (MCP) segments, FUVA and FUVB, separated by an 9 mm gap, and a cross delay line anode. A microchannel plate consists of a slab of highly resistive material with a dense array of tiny tubes (microchannels), each of which is a continuous electron multiplier. The MCPs have an opaque pho-

[2]The groove efficiency is simply a measure of the proportion of diffracted light concentrated into a given order at a particular wavelength. Higher efficiencies are obtained by controlling the cross-sectional profile of the grooves in a diffraction grating using a technique called blazing. A triangular profile is the one most commonly used.

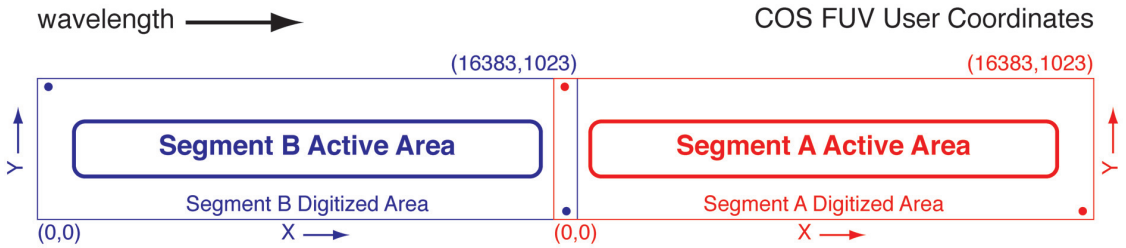

Figure 2.3: A schematic of the COS FUV detector, sourced from the COS Instrument Handbook (Debes et al., 2015).

tocathode deposited on the surface, and a fine wire mesh suspended above this to increase the quantum efficiency. There is also an ion-repeller grid that prevents low energy thermal ions from entering the detector, which reduces the background count rate. Each MCP segment is a 'Z stack', meaning that it consists of three panels, each with microchannels angled away from the normal so as to create a 'Z' shape. In this configuration, the stacks operate at a gain of $\sim 10^7$ electrons per photon impact event (McPhate et al., 2000). The electron cloud produced by each event is accelerated by an applied electric field onto the cross delay line anode. This anode determines the position of each photon impact by splitting the incident electron cloud into four components, each of which travels down a 'delay line'. The timing of the output signals from the delay lines gives the location of the incident electron cloud along both the dispersion (x) and cross-dispersion (y) axis, and hence a location for the initial photon impact (see McPhate et al., 2000, for more details). The locations of the detected photons are then digitised and placed into two arrays (one each for FUVA and FUVB), each with 16 384 $\times$ 1024 pixels, although the active area of the detector is somewhat smaller than this. Each pixel spans 6 $\times$ 24μm. A schematic of the detector is presented in Figure 2.3.

Five different central wavelength settings for both the G130M and G160M gratings are supported to enable full wavelength coverage across the MCP segment gap in various configurations (see table 5.3 in the COS Instrument Handbook; Debes et al., 2015).[3] The large size of the detector means that most of the FUV bandpass may be covered with just four exposures. This is hugely beneficial to

[3]Available online at http://www.stsci.edu/hst/cos/documents/handbooks/current/cos_cover.html.

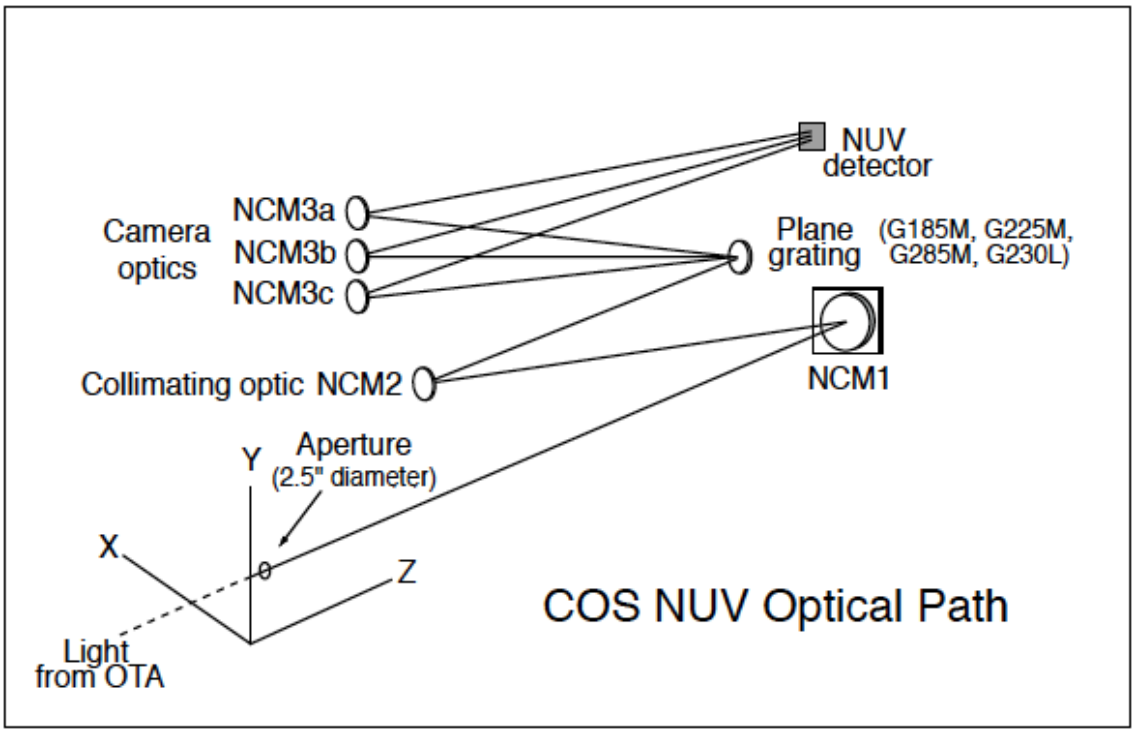

Figure 2.4: The COS NUV optical path, sourced from the COS Instrument Handbook (Debes et al., 2015).

QSO absorption line studies, since one cannot know a priori where absorption features of interest will be detected along the dispersion axis of the detector.

2.2.2 NUV design

The NUV channel was not originally planned by the COS development team, but was added at the request of NASA shortly after its selection (Green et al., 2012). It is intended as a backup facility to the already existing NUV channel on STIS, and was designed to be low cost, so as not to hinder the development of the FUV channel. As a result, three far smaller and cheaper multi-anode microchannel array (MAMA) detectors are used for the NUV channels, labelled NUVA, NUVB and NUVC. These detectors work in a very similar way to the FUV detectors. First order holographic gratings are used, for which the detectors can only provide limited wavelength coverage in a single setting. Many settings are therefore required to cover the entire NUV bandpass, which are selected via a scanning mechanism. The available settings, and the wavelength regions they cover, are listed in table

5.4 of Debes et al. (2015). The introduction of a scanning mechanism for the NUV modes meant that the gratings could not be used to correct for optical aberrations, since they are a function of optical geometry, which changes upon selection of different wavelength regions in this way. As a result, additional optical elements had to be incorporated into the NUV design. Flat gratings are used, with uniform straight grooves, as well as three mirrors that correct, collimate and re-image the light. These mirrors are labelled in Figure 2.4, which presents the optical layout of the NUV channel.

2.2.3 Flight performance

All aspects of the original COS design specifications were met during in-flight testing except for the spectral resolution of the FUV gratings. In fact, the throughput in the FUV, and the SNR compared to that of STIS in the same integration time (aimed to be greater by a factor of 10), greatly exceeded design specifications (Green et al., 2012). The spectral resolution of the COS FUV 'M' gratings were originally designed to be $R \geq 20\,000$ at all wavelengths, but the on-orbit performance is below this level, with the majority of the wavelength coverage falling below $R = 20\,000$. The NUV channels meet pre-flight expectations, but the spectral resolution of the FUV channels is degraded due to the wavefront errors present in the *HST* primary and secondary mirrors (Ghavamian et al., 2009), and the roughness of the primary mirror at the micron level (Kriss, 2011). These errors result in a non-Gaussian line-spread function (LSF), with power moved from the core of the profile to the wings. This effect is more prominent at shorter wavelengths, and is the primary reason for the loss in spectral resolution. Absorption profiles in COS FUV data therefore have depths shallower than one would obtain if the LSF was approximately Gaussian, and any line fitting must take into account the appropriate LSF in order to model absorption line profiles correctly. The LSF of the NUV gratings also differs from a Gaussian, but to a lesser extent.

The COS FUV LSF at 3000 Å is shown in Figure 2.5, in comparison to a Gaussian obtained from ground-based vacuum testing. Wavefront errors introduced from the primary and secondary mirrors on *HST* are the dominant cause for the

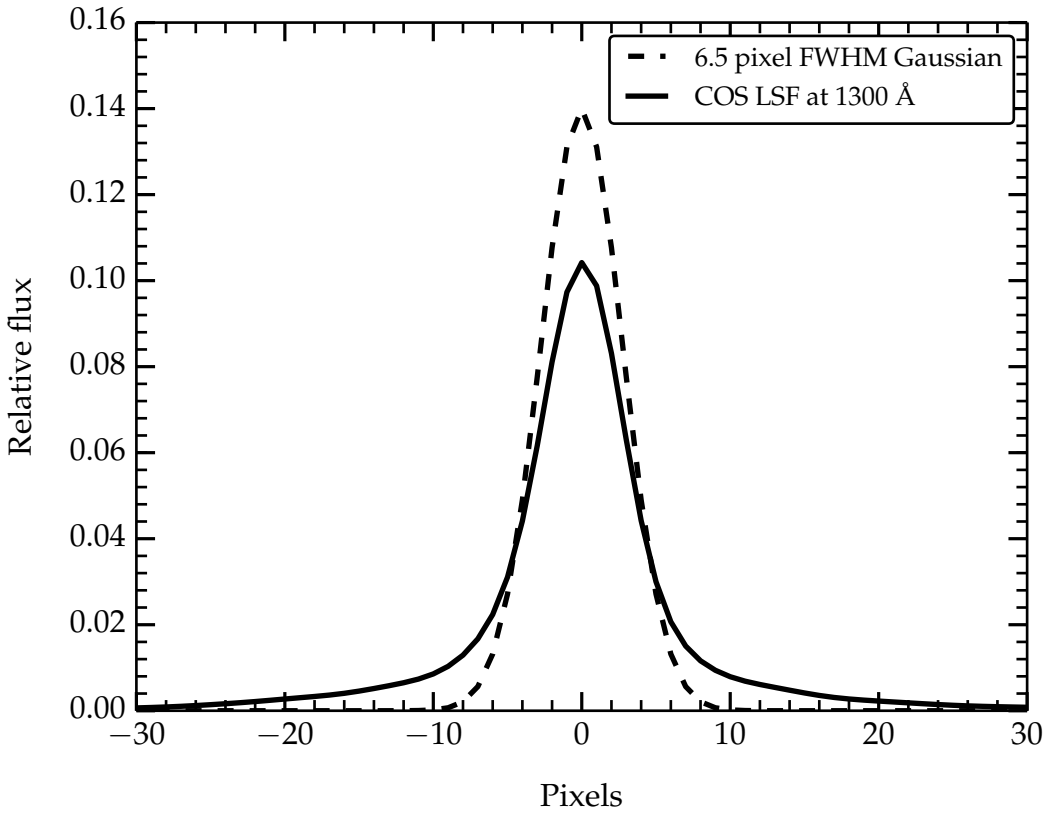

Figure 2.5: A comparison between a Gaussian LSF with FWHM = 6.5 pixels (dashed line), consistent with the expectation from ground-based vacuum testing, and a model LSF profile for the COS FUV G130M grating at 1300 Å (solid line) that includes both the effects of wavefront errors in the *HST* primary and secondary mirrors, and micron-level 'roughness' in the *HST* primary mirror.

shape of the FUV LSF, which clearly deviates from a Gaussian, having far more power in the wings of the profile. This makes the detection of narrow ($b < 35$ km s^{-1}) absorption features more difficult, and also impacts the ability to resolve closely separated absorption components (see Ghavamian et al., 2009; Kriss, 2011, for more details).

Further to the loss in spectral resolution, the other main flight performance concern for COS is the phenomenon of 'gain sag', whereby the FUV detectors start to become less efficient at turning photon impacts into the production of electrons after prolonged exposure to light. The total charge carried by an electron cloud (referred to as the 'pulse height') generated by an event determines whether or not that event is considered a real photon impact or a background event via two

threshold values (one high, and one low). When the effect of gain sag reduces the photon-to-electron conversion efficiency to the point where real photon impacts occasionally produce pulse heights below the low threshold value, a loss in detector sensitivity starts to occur. In extreme cases, this can result in the appearance of spurious absorption features in COS FUV spectra. Gain sag initially affects regions of the detector exposed to bright geocoronal lines, but eventually affects the entire detector.

To mitigate the effects of gain sag, the Space Telescope Science Institute (STScI) has implemented a number of different strategies. Some of these enter the pipeline reduction software, CALCOS. In particular, the lower pulse height threshold value in CALCOS has been reduced over the course of time, at the expense of increasing the detector background. In addition, low gain pixels are flagged, and may be de-weighted or excluded by co-adding spectra from a number of grating positions, offset in wavelength from the nominal values (known as 'FP-POS' positions). Changes have also been made to COS itself, one such example being an increase in the voltage applied across the MCPs, so as to increase the detector gain. Ultimately however, regions of the FUV detectors are becoming irreparably worn out, so the position at which the object spectrum lands on the detectors has to be moved. This occurred on 23 July 2012, when the COS FUV spectral position was moved to 'lifetime position 2', and again on 9 February 2015, with a move to 'lifetime position 3'. Further details on the effects of gain sag on the COS FUV detectors can be found in the COS Instrument Handbook (Debes et al., 2015).

Properties of the various spectroscopic modes derived from on-orbit performance are summarised in Table 2.1.

2.2.4 Noise properties

The noise properties of the COS FUV modes have been analysed in detail by Keeney et al. (2013). The key finding is that the noise in the FUV data is not completely random. For random noise, the measured relationship between the SNR

Table 2.1: COS spectroscopic modes, sourced from the COS Instrument Handbook (Debes et al., 2015).

Grating	Wavelength range (Å)	Bandpass per exposure and FUV gap[a](Å)	Inferred PSA resolving power $R = \lambda/\mathrm{FWHM}$[b]	Dispersion (mÅ pixel^{-1})
		FUV channel		
G130M	900 – 1236	295 / 16	up to 11 500[c]	9.97
	1065 – 1365	296 / 15.7	10 000 – 15 000[c]	9.97
	1150 – 1450	292 / 14.3	16 000 – 21 000	9.97
G160M	1405 – 1775	360 / 18.1	16 000 – 21 000	12.23
G140L	1230 – 2050	> 1150 / 112	1500 – 4000	80.3
		NUV channel		
G185M	1700 – 2100	3 × 35	16 000 – 20 000	37
G225M	2100 – 2500	3 × 35	20 000 – 24 000	33
G285M	2500 – 3200	3 × 41	20 000 – 24 000	40
G230L	1650 – 3200[d]	(1 or 2) × 398	2100 – 3900	390

[a] Width of the gap between FUV detector segments.
[b] The full width at half maximum (FWHM) is that of the empirically determined LSF, which is not Gaussian. R increases approximately linearly with wavelength, unless otherwise stated.
[c] R falls with increasing wavelength. $R \sim 8500 - 11\,500$ between 940 and 1080 Å.
[d] Some shorter wavelengths are recorded in second order light.

in a spectrum smoothed over x pixels, $(\mathrm{S/N})_x$, and the SNR per pixel, $(\mathrm{S/N})_1$, is

$$\eta(x) \equiv \frac{(\mathrm{S/N})_x}{(\mathrm{S/N})_1} = x^{1/2}. \tag{2.2.1}$$

However, the measured values of η deviate from this expectation in individual FUV exposures, and they deviate to an even greater degree in the co-added spectra. The co-addition process involves interpolating individual exposures onto a common linear wavelength scale. Interestingly, the choice of interpolation scheme produces varying levels of deviation from the Poissonian noise expectation,[4] and it is found that a nearest-neighbour interpolation scheme minimises this deviation in the co-added spectra.

The source of non-random noise in the COS FUV data is likely attributable to the design of the FUV detectors. In particular, shadows are produced by the grid wires used to improve the quantum efficiency, and there are irregularities in the

[4]The COS detectors are photon counters, which means they have no read-out noise and their gain is unity.

digitised spectra introduced by fixed hexagonal and moiré patterns in the MCPs, which are not completely corrected for by the CALCOS flat-field. There are also the effects of gain sag. Co-added spectra produced from multiple exposures with different FP-POS positions are typically employed to dither around grid wire shadows and regions affected by gain sag, however this introduces small differences in the fixed hexagonal and moiré patterns from the MCPs along the cross-dispersion direction. These differences do not average out in general, and can constructively interfere, which may explain the amplified deviations from Poissonian noise in the co-added data. For further details, see Keeney et al. (2013).

2.3 The QSO sample

In this thesis, we make use of *HST*/COS and FOS spectroscopy of 60 QSOs to characterise the IGM through observations of intervening absorption systems. Of these, 7 have COS data that were obtained by our collaboration (*HST* general observer (GO) programmes 12264 and 11585), and are presented in published work (Tejos et al., 2014; Finn et al., 2014). The remaining 53 come from a variety of scientific programmes, and are presented by different authors. The full sample of QSOs is summarised in Table 2.2, where we also list the name of the field for which each QSO is associated. For each of these fields we have a spectroscopic catalogue of galaxies constructed from a variety of different instruments and surveys. For a complete description of these galaxy surveys, see Chapter 4.

Data from GO programme 12264 were obtained with the goal of investigating the statistical relationship between intergalactic absorbers and galaxies at $z < 1$. Four QSOs were selected at $z > 1$ for this purpose, namely FBQS J0209−0438, HE 1003+0149, SDSS J135726.27+043541.4 and FBQS J2218+0052. These QSOs were favoured for lying in fields that were already well surveyed for their galaxy content, as well as being sufficiently bright in the FUV and NUV such that a SNR of ~ 10 was achievable with COS in a relatively modest integration time. Specifically, FBQS J0209−0438 lies in the Very Large Telescope (VLT) Visible Multi-Object Spectrograph (VIMOS) Public Extragalactic Redshift Survey (VIPERS) W1 field (Guzzo

Table 2.2: The QSO sample

QSO name	Field name	$\alpha_{\rm QSO}^{\rm J2000}$ ($^{\rm hms}$)	$\delta_{\rm QSO}^{\rm J2000}$ ($^\circ\,'\,''$)	$z_{\rm QSO}$
PG 0003+158	J0005+1609	00:05:59.24	+16:09:49.00	0.451
PG 0026+129	J0029+1609	00:29:13.60	+13:16:04.01	0.142
HE 0056−3622	J0058−3606	00:58:37.36	−36:06:04.79	0.164
LBQS 0107−0235A	J0110−0218	01:10:13.19	−02:19:53.72	0.960
LBQS 0107−0235B	J0110−0218	01:10:16.30	−02:18:51.62	0.956
LBQS 0107−0232	J0110−0218	01:10:14.47	−02:16:58.40	0.726
B0117−2837	J0120−2821	01:19:35.70	−28:21:31.43	0.349
Ton S210	J0120−2821	01:21:51.50	−28:20:57.98	0.116
PG 0157+001	J0159+0023	01:59:50.25	+00:23:40.85	0.163
FBQS J0209−0438	J0209−0438	02:09:30.74	−04:38:26.30	1.132
HE 0226−4110	J0228−1904	02:28:15.17	−40:57:14.29	0.493
PKS 0405−123	J0407−1211	04:07:48.50	−12:11:37.00	0.574
RBS 542	J0426−5712	04:26:00.72	−57:12:00.97	0.104
PKS 0558−504	J0559−5026	05:59:47:30	−50:26:52.01	0.137
SDSS J080908.13+461925.6	J0809+4619	08:09:08.13	+46:19:25.54	0.656
PG 0832+251	J0835+2459	08:35:35.81	+24:59:40.20	0.330
PG 0844+349	J0847+3445	08:47:42.47	+34:45:04.39	0.064
Mrk 106	J0919+5521	09:19:55.34	+55:21:37.08	0.123
RXS J09565−0452	J0956−0453	09:56:30.13	−04:53:17.09	0.155
PG 0953+414	J0956+4115	09:56:52.39	+41:15:22.25	0.234
PG 1001+291	J1007+2929	10:04:02.61	+28:55:35.40	0.330
HE 1003+0149	J1005+0134	10:05:35.25	+01:34:46.13	1.081
FBQS J1010+3003	J1007+2929	10:10:00.69	+30:03:21.56	0.256
Ton 1187	J1013+3551	10:13:03.18	+35:51:23.76	0.079
PG 1011−040	J1014−0418	10:14:20.68	−04:18:40.28	0.058
LBQS 1019+0147	J1022+0132	10:22:18.99	+01:32:18.82	0.789
1ES 1028+511	J1031+5052	10:31:18.52	+50:53:35.88	0.360
1SAX J1032.3+5051	J1031+5052	10:32:16.14	+50:51:19.69	0.173
PG 1048+342	J1058+3412	10:51:43.90	+33:59:26.70	0.167
PG 1049−005	J1051−0051	10:51:51.44	−00:51:17.68	0.360
HS 1102+3441	J1058+3412	11:05:39.82	+34:25:34.64	0.509
SBS 1108+560	J1118+5728	11:11:32.17	+55:47:26.09	0.767
PG 1115+407	J1121+4113	11:18:30.29	+40:25:54.01	0.155
PG 1116+215	J1119+2119	11:19:08.68	+21:19:18.01	0.176
PG 1121+422	J1121+4113	11:24:39.18	+42:01:45.01	0.225
SBS 1122+594	J1118+5728	11:25:53.79	+59:10:21.58	0.852
Ton 580	J1131+3114	11:31:09.48	+31:14:05.50	0.290
3C 263	J1139+6547	11:39:57.02	+65:47:49.42	0.646
PG 1216+069	J1226+0319	12:19:20.93	+06:38:38.51	0.331
3C 273	J1226+0319	12:29:06.70	+02:03:08.60	0.158
HE 1128+0131	J1226+0319	12:30:50.04	+01:15:22.68	0.117
PG 1229+204	J1232+2009	12:32:03.63	+20:09:29.56	0.063
PG 1259+593	J1301+5902	13:01:12.93	+59:02:06.76	0.478
PKS 1302−102	J1305−1033	13:05:33.00	−10:33:18.00	0.278
PG 1307+085	J1309+0819	13:09:47.00	+08:19:48.25	0.155
SDSS J135726.27+043541.4	J1357+0435	13:57:26.28	+04:35:41.50	1.232
PG 1424+240	J1427+2348	14:27:00.40	+23:48:00.04	0.604
PG 1435−067	J1438−0658	14:38:16.16	−06:58:20.50	0.126
LBQS 1435−0134	J1437−0147	14:37:48.28	−01:47:10.79	1.310
Mrk 478	J1442+3526	14:42:07.47	+35:26:22.96	0.079
Ton 236	J1528+2825	15:28:40.61	+28:25:29.86	0.450

Table 2.2: continued

QSO name	Field name	$\alpha_{\rm QSO}^{\rm J2000}$ ($^{\rm hms}$)	$\delta_{\rm QSO}^{\rm J2000}$ ($^\circ\,'\,''$)	$z_{\rm QSO}$
1ES 1553+113	J1555+1111	15:55:43.04	+11:11:24.32	0.414
Mrk 877	J1620+1724	16:20:11.28	+17:24:27.58	0.112
PKS 2005−489	J2009−4849	20:09:25.30	−48:49:53.00	0.071
Mrk 1513	J2132+1008	21:32:27.82	+10:08:19.18	0.063
PHL 1811	J2155−0922	21:55:01.50	−09:22:25.00	0.192
PKS 2155−304	J2158−3013	21:58:52.10	−30:13:31.01	0.117
FBQS J2218+0052	J2218+0052	22:18:06.67	+00:52:23.63	1.273
MR 2251−178	J2254−1734	22:54:05.80	−17:34:54.98	0.064
4C 01.61	J2351−0109	23:51:56.12	−01:09:13.36	0.174

et al., 2014), and HE 1003+0149, SDSS J135726.27+043541.4 and FBQS J2218+0052 lie in the VLT VIMOS Deep Survey (VVDS) F10, F14 and F22 fields respectively (Le Fevre et al., 2005). We note that FBQS J0209−0438 was originally selected on the virtue of intersecting the Gemini Deep Deep Survey (GDDS) (Abraham et al., 2004), however the more recent VIPERS supersedes this survey vastly in terms of galaxy completeness and survey depth.

Data from GO programme 11585 were obtained with the goal of measuring the characteristic sizes of H I absorption systems, and investigating their connection with galaxies at $z < 1$. For this purpose, three QSOs at $z \sim 1$ were selected that have a very close angular separation on the sky, such that the transverse distances separating their sight-lines over the total redshift path-length probed spans the range $\sim 0.01 - 2$ Mpc. These are LBQS 0107−0235A, LBQS 0107−0235B and LBQS 0107−0232. This is the only known FUV-bright triplet that probes such small transverse scales. These data were obtained from the FUV modes of COS, with NUV data already being available from *HST*/FOS. The FOS data on this QSO triplet, along with a modest galaxy survey around these sightlines, were already analysed and presented in a paper by our collaboration (Crighton et al., 2010). Since then, a much more comprehensive, deep galaxy redshift survey to $z \sim 1$ has been constructed around these sight-lines, making use of the Deep Imaging Multi-Object Spectrograph (DEIMOS), Gemini Multi-Object Spectrograph (GMOS) and VIMOS (see Tejos et al., 2014, and Chapter 4 for a complete desription).

The coincidence of the aforementioned QSOs with these galaxy redshift surveys allows for an investigation into the galaxy-IGM connection up to $z \approx 1$, which

extends far beyond that available to shallower galaxy surveys such as the Sloan Digital Sky Survey (SDSS), 2dF Galaxy Redshift Survey (2dFGRS) and Galaxy and Mass Assembly (GAMA) survey. Nevertheless, a large number of QSOs observed with COS overlap with these and other low redshift surveys, for which a thorough analysis has been performed by the COS guaranteed time observations (GTO) team (Danforth et al., 2014). We therefore include this sample of QSOs in addition to those analysed by our collaboration to increase the statistical power of the overall sample.

Details on the QSO observations are listed in Table 2.3, along with the proposal IDs for those observations. The proposal abstracts relating to all of the QSOs listed may be obtained by searching for that ID on the Mikulski Archive for Space Telescopes (MAST).[5] The QSOs in Table 2.3 were selected for a variety of different science cases, however we note that this does not bias any results we obtain by studying the IGM-galaxy connection *foreground* to those QSOs, as these foreground regions are essentially random for the majority of the science cases.

In the following sections, we describe the data reduction and analysis procedures performed on our COS data set obtained from the GO programmes 12264 and 11585, and also on the QSO LBQS 1435−0134 obtained from GO programme 11741. For all the other QSOs in our sample, the reduction and analysis procedures are described in Danforth et al. (2014), and for these we downloaded the high-level science products available on the MAST archive.[6]

2.4 COS data reduction

COS data direct from *HST* are processed through the STScI Operational Pipeline Unified System (OPUS). This pipeline performs a simple conversion, unpacking data from individual exposures on COS, and combining them into files containing raw, uncalibrated data. These files are then downloaded from the MAST archive, and the data reduced using the COS pipeline, CALCOS. A complete description of

[5] http://archive.stsci.edu/hst/search.php
[6] http://archive.stsci.edu/prepds/igm/

Table 2.3: The QSO observations

QSO name	G130M S/N[b]	G160M S/N[b]	NUV[a] S/N[b]	Programme ID(s)
PG 0003+158	22	19	–	12038
PG 0026+129	18	–	–	12569
HE 0056−3622	25	15	–	12604
LBQS 0107−0235A	9	8	30	11585, 6592, 6100, 5320
LBQS 0107−0235B	9	7	30	11585, 6592, 6100, 5320
LBQS 0107−0232	–	7	18	11585, 6592, 6100
B0117−2837	24	19	–	12204
Ton S210	41	26	–	12204
PG 0157+001	16	–	–	12569
FBQS J0209−0438	12	10	12	12264
HE 0226−4110	34	24	–	11541
PKS 0405−123	59	30	–	11508, 11541
RBS 542	61	35	–	11686
PKS 0558−504	19	10	–	11692
SDSS J080908.13+461925.6	15	13	–	12248
PG 0832+251	14	12	–	12025
PG 0844+349	18	–	–	12569
Mrk 106	28	18	–	12029
RXS J09565−0452	16	–	–	12275
PG 0953+414	38	26	–	12038
PG 1001+291	21	17	–	12038
HE 1003+0149	9	9	–	12264
FBQS J1010+3003	17	10	–	12025
Ton 1187	16	–	–	12275
PG 1011−040	29	18	–	11524
LBQS 1019+0147	6	5	–	11598
1ES 1028+511	20	13	–	12025
1SAX J1032.3+5051	12	6	–	12025
PG 1048+342	23	16	–	12024
PG 1049−005	14	12	–	12248
HS 1102+3441	17	13	–	11541
SBS 1108+560	4	14	–	12025
PG 1115+407	23	15	–	11519
PG 1116+215	39	28	–	12038
PG 1121+422	21	13	–	12604
SBS 1122+594	14	12	–	11520
Ton 580	21	16	–	11519
3C 263	34	23	–	11541
PG 1216+069	24	16	–	12025
3C 273	73	–	–	12038
HE 1128+0131	44	36	–	11686
PG 1229+204	17	–	–	12569
PG 1259+593	32	24	–	11541
PKS 1302−102	26	20	–	12038
PG 1307+085	18	–	–	12569
SDSS J135726.27+043541.4	9	7	11	12264
PG 1424+240	21	21	–	12612
PG 1435−067	14	–	–	12569
LBQS 1435−0134	23	16	–	11741
Mrk 478	18	–	–	12569
Ton 236	18	15	–	12038

Table 2.3: continued

QSO name	G130M S/N[b]	G160M S/N[b]	NUV[a] S/N[b]	Programme ID(s)
1ES 1553+113	33	26	–	11520, 12025
Mrk 877	18	–	–	12569
PKS 2005−489	24	15	–	11520
Mrk 1513	32	20	–	11524
PHL 1811	36	24	–	12038
PKS 2155−304	45	–	–	12038
FBQS J2218+0052	–	–	10	12264
MR 2251−178	38	30	–	12029
4C 01.61	20	–	–	12569

[a] FOS gratings G270H and/or G190H for LBQS 0107−0235A, LBQS 0107−0235B and LBQS 0107−0232, COS G230L grating otherwise.
[b] Median SNR per resolution element.

CALCOS is beyond the scope of this thesis, however a few modifications are made to the default CALCOS reduction parameters, and we describe them here.

CALCOS implements a simple boxcar extraction scheme to obtain one-dimensional spectra from the flat-field corrected images. We optimise this scheme by narrowing the source extraction box from its default size to match the apparent size of any given source in the cross-dispersion direction upon inspection of each flat-fielded image. Background extraction boxes are also enlarged to encompass more of the background signal, making sure to avoid the BOA and regions close to the detector edges. By default, CALCOS performs a boxcar smoothing on the background counts at each pixel along the dispersion axis to provide a robust measure of the background. This background smoothing is applied everywhere, including areas affected by scattered light from strong geocoronal emission lines. Particularly for optimised extraction windows, this leads to an overestimation of the background in these regions and their immediate vicinity due to 'smearing' of the light. To avoid this, we set the background smoothing length in CALCOS to 1 pixel and perform our own background smoothing on the extracted spectra in post-processing. This procedure masks out affected portions of the spectrum, namely the Lyα and O I $\lambda\lambda$1302 + 1306 geocoronal emission lines, then interpolates across the gap to estimate the actual background level in these regions.

In Figure 2.6 we show an example spectrum, that of the QSO 3C 273, extracted with CALCOS. In the top panel is the flat-fielded image, with the object spectrum

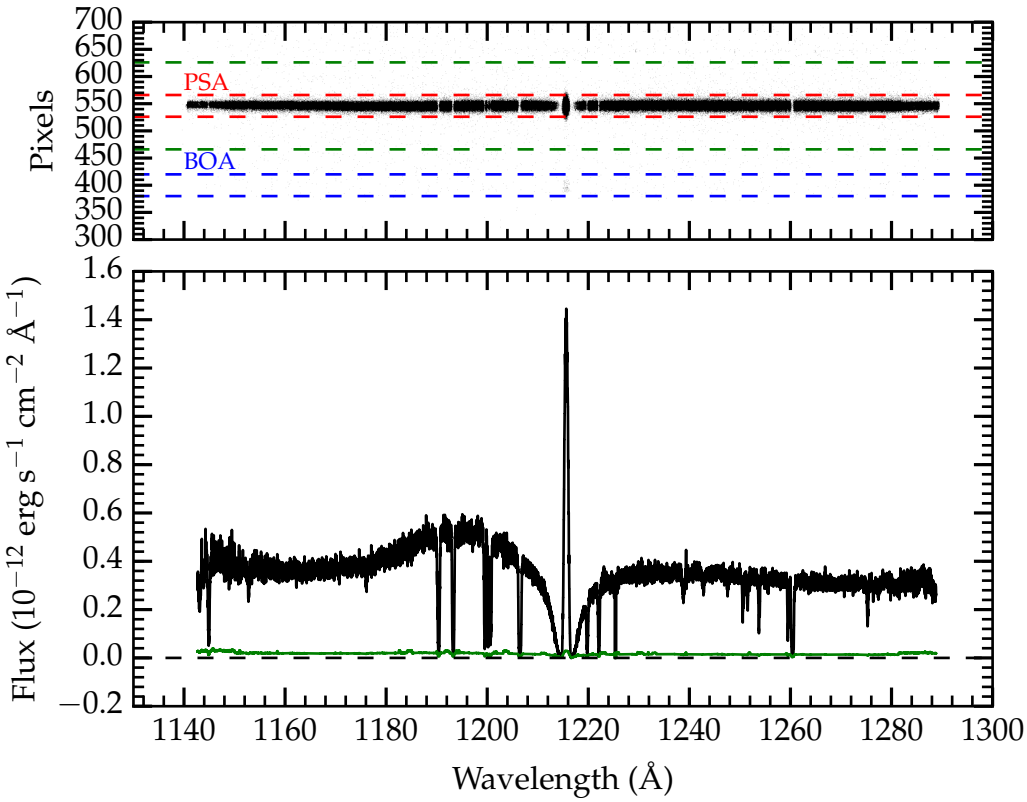

Figure 2.6: The COS G130M FUVB spectrum of 3C 273. The top panel shows the flat-field corrected 2D spectrum, with the effective location of the PSA and BOA labelled, and the background extraction window indicated in green. The bottom panel shows the flux calibrated 1D spectrum in black, and the 1σ statistical flux uncertainty in green. The zero-level is indicated by the dashed black line. The Lyα geocoronal line is clearly seen at $\approx$ 1215 Å.

clearly visible. The position and extent of the primary science aperture (PSA) and bright object aperture (BOA) is shown, along with the background extraction windows. As is highly evident from the figure, the background level for COS is extremely small. A large smoothing length must be applied in order to obtain a robust estimate of the background level. Smoothing lengths are chosen to be as large as possible, whilst still preserving background features in the dispersion direction. Actual values vary from spectrum to spectrum, since the background level varies as a function of time, but they are typically in the range of 100 – 1000 pixels, and are chosen based on visual inspection. The bottom panel shows the extracted and flux calibrated 1D spectrum, along with the statistical flux uncer-

tainty (proportional to $\sqrt{n}$, where n is the number of photon counts) and the zero level indicated. The Lyα geocoronal line is particularly prominent, and can also be seen in the 2D spectrum. In the latter, it is clear that the line is broader in the cross-dispersion direction than the rest of the spectrum, and given the adopted source extraction window in red, there is clear justification for the modified background extraction procedure described.

After extracting 1D spectra using CALCOS, individual exposures from multiple settings are aligned and co-added using custom-built PYTHON routines developed by the author.[7] These are loosely based upon IDL routines developed by the COS GTO team.[8] They work as follows:

Individual `x1d` files (`*_x1d.fits`) produced in the CALCOS reduction are collated, along with their header information. Each of these files corresponds to a one-dimensional extracted spectrum from a single central wavelength setting and FP-POS position. FUV files contain two data extensions corresponding to FUVA and FUVB. NUV files contain three extensions corresponding to NUVA, NUVB and NUVC. Data quality flags[9] for all pixels are assigned new flags to mean one of three options: (i) retain pixel for co-addition with a weight equal to 1, (ii) retain pixel for co-addition with a weight equal to 0.5, or (iii) discard pixel from the co-addition process. Flags assigned with option (i) are those where no anomalous condition is noted, or where some unusual features have been identified in long background exposures. The latter is not expected to have any effect on the final data products except perhaps where the count rate is very low or the background higher than normal. Flags assigned with option (ii) are those in regions where the background count rate is apparently higher than the surrounding region and/or is unstable, and in regions on the FUV detector where the gain is low enough so as to affect the calibration (gain sag). For our observations, we made use of all the available FP-POS positions, so the affect of gain sag should be minimal having effectively dithered around these features, which are fixed on the detector plane.

[7] https://github.com/cwfinn/COS/
[8] http://casa.colorado.edu/~danforth/science/cos/costools.html
[9] http://www.stsci.edu/hst/cos/pipeline/cos_dq_flags

For the NUV channel only, regions affected by detector shadows (vignetting in this case) are assigned option (ii). These vignetted areas affect significant portions of the NUV spectra. and so are retained to avoid large gaps in the final co-added spectra. A weight of 0.5 ensures that the data in these regions contributes significantly towards increasing the SNR, whilst minimising any additional error in the flux calibration. All other flags are assigned option (iii). Data quality flags that are assigned either of the first two options are referred to as 'good', and those assigned the latter option are referred to as 'bad.'

Next, the background counts are re-estimated as was previously motivated, boxcar smoothing only across pixels not affected by scattered geocoronal light and with good data quality flags. The error array from CALCOS is then re-calculated using the new background smoothing lengths and each spectrum is flux calibrated using the time dependent sensitivity curves provided by STScI, corrected to the epoch of observation.

Exposures are now co-aligned by cross-correlating regions centred on strong Galactic absorption features. Specifically, these are C II λ1334, Al II λ1670, Si II λ1260, Si II λ1526 and Mg II $\lambda\lambda$2796, 2803. Using these features allows for co-alignment between all settings in the FUV gratings (assuming the FUV wavelength scale from CALCOS is relatively correct), and some NUV settings. For each grating, we pick the central wavelength setting and FP-POS position with the most accurately determined wavelength solutions from STScI as a reference. We assume that CALCOS correctly shifts these configurations into a heliocentric reference frame. All other settings for each grating are then cross-correlated with these ones if the reference and comparison settings both contain one of the absorption features specified. Wavelength offsets are then applied to the comparison settings to match the reference ones. These offsets typically amount to a resolution element or less. For those settings that cannot be aligned on any of the Galactic features specified, we search for other strong absorption lines on which to perform the cross-correlation. Once each exposure has been aligned, they are then scaled so that their median flux values match in overlap regions.

Before performing the final co-addition of the data, flux and error values as-

signed to pixels with bad data quality flags are set to zero, and pixels flagged for de-weighting have their exposure time halved. Following Keeney et al. (2012), flux and error values are then placed on a linear wavelength scale using a nearest-neighbour interpolation to minimise the effects of non-Poissonian noise on the co-added data products. For any point along this scale, the linear spacing is set to the dispersion of the grating that provides the spectral coverage in that region. The co-addition is then performed using modified exposure-time weighting, i.e. flux values are co-added according to the following rule:

$$F_i = \frac{\sum_j F_j \times t_{\mathrm{exp};\, j}}{\sum_j t_{\mathrm{exp};\, j}}, \tag{2.4.1}$$

where i represents the ith pixel in the final, co-added spectrum, and j represents the jth pixel that is co-added to make pixel i. Similarly, error values are co-added as follows:

$$\delta F_i = \frac{\sqrt{\sum_j (\delta F_j \times t_{\mathrm{exp};\, j})^2}}{\sum_j t_{\mathrm{exp};\, j}}. \tag{2.4.2}$$

Finally, the combined spectra are re-binned to ensure Nyquist sampling, i.e. two pixels per resolution element.

In the sections that follow, we describe the analysis procedures performed on the reduced, co-added data.

2.5 Continuum fitting

The first step in constructing an absorption line list from a QSO spectrum is to obtain an estimate of the pseudo continuum (i.e. the continuum including emission lines), such that we may calculate the transmission of UV photons from the QSO through the IGM. To do this, we use a technique similar to that described in Young et al. (1979), Carswell et al. (1982) and Aguirre et al. (2002). This involves fitting an arbitrary number of polynomial segments (a polynomial spline) to the observed continuum emission. Continuum fitting is somewhat uncertain due to

the fact that regions of the continuum emission can be completely lost to multiple blended absorption features arising from the Lyα forest. Luckily, at $z < 1$, the instances of multiple blended absorption features are few and far between, and this only becomes a big problem at higher redshifts. Ideally, one would fit a physical model for the observed continuum emission, but a full theoretical understanding on the spectral energy distributions (SEDs) of QSOs is far from complete. We are therefore resigned to estimating the continuum in an inherently phenomenological way.

To fit the QSO continuum, the spectrum is first divided up into an arbitrary number of segments in wavelength. These segments are on average around 12 Å wide through the Lyα forest (blue-wards of the Lyα emission line), and narrower in regions containing prominent emission lines (where the continuum varies most significantly). Larger segments are used outside the Lyα forest in regions free of emission lines. An iterative, three stage process is then employed as follows:

1. The median flux value is calculated for each segment.

2. A first order spline (a series of straight line segments that have matching y-values at their boundaries) is fitted through the set of points defined by the central wavelength and the median flux value of each segment.

3. Flux values falling a specified number of standard deviations ($n\sigma$) below the fitted spline are removed.

This process is repeated until the continuum converges upon a solution. As a final step, a cubic spline (a series of 3rd order polynomials that have matching zeroth, first and second order derivatives at their boundaries) is fit through the set of median flux values, so that a smooth continuum is achieved. The number of spline segments, and the width of each is arbitrary, however choosing too few typically results in a continuum that does not have enough small scale variation to capture the data, and too many results in a continuum that is too strongly influenced by absorption lines. So long as the absorption line density is low enough not to destroy observable features in the continuum, a sensible choice for the number of

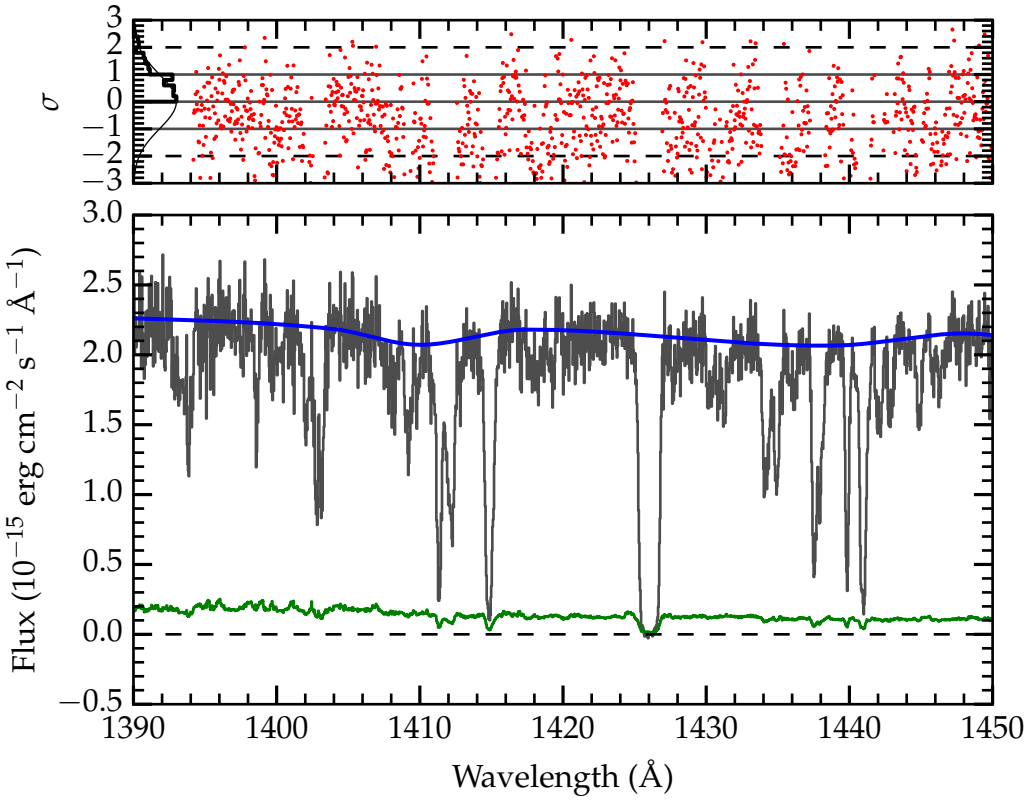

Figure 2.7: An example continuum fit to a region of the spectrum of FBQS J0209−0438. The data is shown in grey in the bottom panel, along with the error in green and the continuum fit in blue. The top panel shows the residuals on the fit, along with a left inset histogram illustrating the distribution of residuals above the continuum, which is roughly Gaussian, indicating a good fit to the data.

spline segments and the $n\sigma$ in step 3 of the process outlined above can produce a reasonably robust continuum estimation.

To determine whether or not a good fit has been achieved, we look to the distribution of residuals above the continuum. In general, a good fit may be indicated by a roughly Gaussian distribution of residuals above the fitted continuum, reflective of the noise in the spectrum. In Figure 2.7, we show the continuum fit to a portion of the spectrum of FBQS J0209−0438. The top panel shows the distribution of residuals above the continuum as a histogram in black, which is clearly well represented by a Gaussian. By trial and error, we are able to pick a number of spline segments, and a value for $n\sigma$, that gives a result like that depicted in Figure 2.7. The best value for $n\sigma$ varies as a function of SNR, spectral resolution, and

location within or outside of the Lyα forest, but typical values are in the range 1.5 – 3σ.

After performing the continuum fit, even after finely tuning the algorithm described above, a small amount of modification by hand is usually required around strongly peaked emission lines, at the spectrum edges, and over Lyman limit systems, where the automated fit performs poorly. For this purpose we make use of an interactive Python/Matplotlib environment, which enables the user to add or remove boundary points between spline segments, thereby modifying the fit (since the spline segments must match at the boundaries). The interactive environment closely resembles the layout of Figure 2.7, so a visual inspection on the quality of the fit is possible. This visual inspection is a crucial element of the fitting process, since, despite being a useful diagnostic, a Gaussian distribution of residuals above the fitted continuum may be achieved even with a poor fit to the data. After fitting the continuum, the entire spectrum is normalised by this continuum estimate to give the overall transmission of UV photons from the QSO through the IGM.

2.6 Absorption line identification

Before fitting Voigt profiles to absorption lines in any QSO spectrum, we first need to obtain a guess as to what atomic element, ionisation stage, and transition each line is attributable to. For this purpose, we have developed an algorithm to allow reliable identification of most of the resolved absorption lines due to the intervening IGM. Identification is performed by inspection of a 'velocity plot' within an interactive Python/Matplotlib environment. This is simply a stack of plots, each representing the expected location of an absorption line attributable to some atomic element (in a particular ionisation stage and due to a particular atomic transition) in velocity space relative to a particular redshift. Within this environment, the user can scan through the spectrum in redshift space from the point-of-view of any number of atomic transitions (based on their rest-frame transition wavelengths). In this way, one can easily identify absorption 'systems' (also referred to

as 'absorbers') at different redshifts by looking for absorption lines due to atomic transitions that should align in velocity space, for example, the Lyman series of transitions attributable to absorption by an intervening structure containing neutral atomic hydrogen (H I).[10]

Due to the high Universal abundance of hydrogen, and the high oscillator strength of the Lyα transition, a large majority of the absorption lines seen in any QSO spectrum will be due to this transition, and indeed a large number from the associated Lyβ and higher order Lyman series transitions. Some metal transitions are seen however, the most common of which are listed in Table 2.4. This table is adapted from that in Morton et al. (1988) to include the most recent data compiled in Verner et al. (1994) and Morton (2003) (see references therein).

The ions in Table 2.4 are compiled using logical arguments, based on those in Morton et al. (1988), for why certain atomic transitions may be more prevalent in QSO spectra than others. It is first assumed that one ionisation stage dominates the populations of a particular element in any particular absorbing structure, and that this element has the solar abundance relative to hydrogen ($N/N_{\rm H}$). In this case, the equivalent width, W_λ(Å), of an unsaturated absorption line with oscillator strength f at wavelength λ(Å) is given by:

$$\log\left(\frac{W_\lambda}{\lambda}\right) = \log\left(\frac{N}{N_{\rm H}}\right) + \log\lambda f + \log N_{\rm H} - 20.053. \tag{2.6.1}$$

Hence, for a detectability limit $W_\lambda \geq 0.015$Å[11] at 1500Å, the column density of hydrogen in all forms $N_{\rm H}$(cm^{-2}) is given by

$$\log N_{\rm H} \geq 15.053 - \log\left(\frac{\lambda f N}{N_{\rm H}}\right). \tag{2.6.2}$$

[10]Here by an absorption 'system', we refer to any set of absorption lines attributable to a single ion at a single velocity. We do not include in this definition adjacent absorption components, which may or may not correspond to complex velocity structure in single physical structures. Our definition therefore does not make any implicit assumptions on the physical nature of the intervening absorbers beyond the fact that they trace some gas moving with some velocity.

[11]This value is approximately the 3σ limiting equivalent width for a SNR $\sim$ 10 spectrum obtained with the COS G130M and G160M gratings.

Table 2.4: A list of the most commonly observed transitions in QSO spectra adapted from that presented in Morton et al. (1988).

Ion	λ (Å)[a]	f-value[a]	$\log(\lambda f) + 12.00 + \log(N/N_{\rm H})$[b]
H I	1215.6701	0.4164	14.704
H I	1025.7223	0.07912	13.909
H I	972.5368	0.029	13.450
H I	949.7431	0.01394	13.122
C II	903.9616	0.007799	11.002
C II	687.0526	0.336	10.883
C II	1334.5323	0.1278	10.752
C II	1036.3367	0.118	10.607
C III	977.0201	0.757	11.389
C IV	1548.2049	0.1899	10.988
C IV	1550.7785	0.09475	10.687
N I	1199.5496	0.132	10.150
N II	915.6131	0.159	10.113
N II	644.6337	0.224	10.110
N III	685.513	0.268	10.214
N IV	765.148	0.616	10.623
N V	1238.821	0.156	10.236
N V[c]	1242.804	0.077	9.931
O I	1302.1685	0.048	10.526
O I	877.8787	0.0589	10.444
O I	988.7734	0.0465	10.393
O I	791.9732	0.0464	10.296
O II	834.4655	0.132	10.772
O II	833.3294	0.0886	10.598
O III	702.332	0.137	10.713
O III	832.927	0.107	10.680
O IV	787.711	0.11	10.668
O IV	608.398	0.067	10.340
O V	629.73	0.515	11.241
O VI	1031.9261	0.1325	10.866
O VI	1037.6167	0.0658	10.564
Mg II	2796.3543	0.6155	10.816
Mg II	2803.5315	0.3058	10.513
Si II	1260.4221	1.18	10.732
Si II	1193.2897	0.582	10.402
Si II	1190.4158	0.292	10.101
Si III	1206.500	1.63	10.854
Si IV	1393.7602	0.513	10.414
Si IV	1402.7729	0.254	10.112
S II	765.693	1.19	10.160
S III	677.746	1.64	10.246
S IV	657.34	1.18	10.090
S V	786.48	1.42	10.248
Fe II	2382.7642	0.32	10.382
Fe II	2600.1725	0.2394	10.294

[a] Vacuum wavelengths and oscillator strengths from Verner et al. (1994) and Morton (2003).

[b] A measure of the transition strength, which takes into account the oscillator strength and the cosmic abundance of the element in question (although not its distribution over ion stages).

[c] Included to ensure both members of the N V doublet are present.

The list is then limited to lines with

$$\log\left(\frac{\lambda f N}{N_{\rm H}}\right) \geq -1.947, \tag{2.6.3}$$

such that these lines are (in theory) observable for any absorbing structure with $\log N_{\rm H} \geq 17.00$ and solar metallicity.[12] We then impose further constraints as follows:

1. Transitions with rest-frame wavelengths outside of the range $600 < \lambda < 2900$ Å are excluded since they are not covered by our COS spectra in any part of the redshift range of interest ($0 < z < 1$).

2. All neutral ions are omitted (as these are not common in QSO spectra, presumably because the ionising radiation field is too strong), except for N I and O I, which both have ionisation potentials higher than that of H I.

3. Excited state fine-structure lines are omitted, as the typical gas densities in the IGM do not favour the formation of these lines.

4. Only the four strongest Lyman series transitions are kept.

The resulting list of atomic transitions, summarised in Table 2.4, is used as an initial search list for the purpose of identifying absorption lines in our QSOs spectra. These are deemed to be the most common intervening absorption lines, but we must also consider absorption lines arising from both gas in our Galaxy, and gas in the QSO host galaxy. The local conditions in these environments are different to those typical in the IGM in general. In particular, the gas densities are expected to be many orders of magnitude higher, and the gas metallicities are also expected, in some cases, to be higher by up to $\sim$ 1 order of magnitude. This favours the production of excited state fine-structure lines, and a larger number of metal transitions. In QSO host galaxies, the local UV radiation field has an intensity much

[12]This limit is imposed somewhat arbitrarily, and the degree to which any particular line is present is also dependent on the metallicity and the local ionization conditions in the gas, which we do not take into account here. Nevertheless, the list is created in an objective way, and reflects in general those lines that are most commonly observed in the spectra of QSOs.

higher than the UV background, which leads to the formation of a number of more highly ionized species than are commonly seen in the IGM, such as Ne VIII and Mg X.

Bearing in mind the considerations above, we also list additional transitions that could potentially be observed in both Galactic (Table 2.5) and QSO host environments (Table 2.6). For the former, we refer to both the Galaxy ISM and halo gas, which is known to contain a population of absorbers moving at velocities offset from the local standard of rest, known as intermediate velocity clouds (IVCs) and high velocity clouds (HVCs) (Wakker & van Woerden, 1997).

The line-identification algorithm then proceeds as follows:

1. Inspect a velocity plot of transitions from both those in Table 2.4 and Table 2.5 that have wavelength coverage from the QSO spectrum centred at $z = 0$, and identify any absorption lines seen at this redshift or ± 500 km s^{-1} away (typical velocities of HVCs; Wakker & van Woerden, 1997) as absorption features due to material in the Milky Way.

2. Inspect a velocity plot of transitions from both those in Table 2.4 and Table 2.6 that have wavelength coverage from the QSO spectrum centred at $z = z_{\rm QSO}$, and identify absorption lines seen at this redshift, or up to -5000 km s^{-1} offset from this, as absorption features due to material in the QSO host galaxy and/or its local environment.[13]

3. Scan through the spectrum with an interactive velocity plot systematically from $z = z_{\rm QSO} - 5000/c$ (where c is the speed of light in km s^{-1}) to $z = 0$, identifying transitions from Table 2.4 that align in velocity, and have the expected ratio of line-strengths. For the identification of metal ions, two transitions from that ion, or at least one accompanying H I transition must be available to confirm the identification. Identification of H I transitions is

[13]These absorbers may or may not show traits typical of QSO associated systems, such as an abundance of highly ionized transitions and equivalent width ratios close to 1:1 (partial covering), but we flag everything to -5000 km s^{-1} offset from the QSO redshift as a conservative assessment of structures not considered intervening.

Table 2.5: Additional absorption lines attributable to the Galaxy ISM and HVCs/IVCs.

Ion	λ (Å)[a]	f-value[a]	$\log(\lambda f) + 12.00 + \log(N/N_{\rm H})$[b]
C I	1656.9284	0.149	10.912
C I	1560.3092	0.0774	10.602
C I	1277.2452	0.0853	10.557
C I	1328.8333	0.0758	10.523
C I	1193.03	0.0409	10.208
C I	1280.1352	0.0263	10.047
C I	1157.9097	0.0212	9.910
C I	1193.9954	0.0124	9.690
C I	1188.8329	0.0124	9.689
C II*	1335.6627	0.01277	9.752
N I[c]	1200.2233	0.0869	9.968
N I[c]	1200.7098	0.0432	9.665
N I[c]	1134.9803	0.0416	9.624
N I[c]	1134.4149	0.0278	9.449
N I[c]	1134.1653	0.0146	9.169
Al II	1670.7886	1.74	9.953
Si II	1526.7070	0.133	9.868
Si II	1304.3702	0.0863	9.611
Si II*	1197.3938	0.145	9.800
P II	1152.818	0.245	8.011
S II	1259.518	0.0166	8.520
S II	1253.805	0.0109	8.336
S II	1250.578	0.00543	8.032
Fe II	1144.9379	0.083	9.478
Fe II	1608.4509	0.0577	9.468
Fe II	1143.226	0.0192	8.841
Fe II	1133.6654	0.00472	8.228
Fe II	1142.3656	0.00401	8.161
Ni II	1370.132	0.056	8.135
Ni II	1317.217	0.057	8.126
Ni II	1741.5531	0.0427	8.121

[a] Vacuum wavelengths and oscillator strengths from Verner et al. (1994) and Morton (2003).

[b] A measure of the transition strength, which takes into account the oscillator strength and the cosmic abundance of the element in question (although not its distribution over ion stages).

[c] Sometimes ignored due to the presence of contaminating N I geocoronal emission.

Table 2.6: Additional absorption lines attributable to the QSO host galaxy.

Ion	λ (Å)[a]	f-value[a]	$\log(\lambda f) + 12.00 + \log(N/N_{\rm H})$[b]
N III	989.799	0.123	10.035
N III*	685.816	0.32	10.291
N III*	991.579	0.11	9.988
O III*	833.742	0.0796	10.552
O III*	702.899	0.0897	10.530
O III**	703.85	0.1386	10.719
O III**	835.292	0.0877	10.595
O IV	554.075	0.224	10.824
O IV*	554.514	0.279	10.920
O IV*	790.199	0.0994	10.625
O IV*	609.829	0.0669	10.341
Ne IV	543.891	0.116	9.860
Ne V	568.42	0.0928	9.782
Ne V*	569.83	0.081	9.724
Ne V**	572.337	0.096	9.800
Ne VI	558.59	0.0907	9.765
Ne VI*	562.805	0.12	9.890
Ne VIII	770.409	0.103	9.960
Ne VIII	780.324	0.0505	9.656
Na IX	681.725	0.0919	8.117
Na IX	694.13	0.045	7.815
Mg X	609.79	0.0842	9.290
Mg X	624.95	0.041	8.989
P IV	950.6569	1.49	8.711
S III	698.731	0.85	9.974
S III*	678.458	0.917	9.994
S III**	680.681	1.38	10.173
S IV	748.397	0.459	9.736
S IV*	661.443	1.201	10.100
S IV*	750.225	0.597	9.851
S VI	933.376	0.436	9.810
S VI	944.525	0.215	9.508
Ar VII	585.748	1.21	9.250
Ar VIII	700.24	0.375	8.819
Ar VIII	713.802	0.18	8.509
Fe III	859.723	0.115	9.495

[a] Vacuum wavelengths and oscillator strengths from Verner et al. (1994), Morton (2003) and the NIST Atomic Spectra Database.

[b] A measure of the transition strength, which takes into account the oscillator strength and the cosmic abundance of the element in question (although not its distribution over ion stages).

set to require at least one accompanying H I transition, or two transitions of a particular metal ion.

4. Scan back through the spectrum (starting at $z = z_{\rm QSO} - 5000/c$), looking in detail at strong H I systems with accompanying metal ions for the presence of weaker metal transitions that aren't listed in Table 2.4.

5. Classify remaining absorption lines (which aren't otherwise constrained and without corresponding Lyβ coverage) to be due to the Lyα transition. Require that these lines have Doppler broadening parameters $b > 10$ km s^{-1} (Lyα lines have typical b-values > 20 km s^{-1}).[14]

The completion of this process results in a list of guessed transitions and their observed wavelengths (and hence redshifts). These, accompanied by initial guesses for the column densities and Doppler broadening parameters, provide the starting point for fitting Voigt profiles, which is the subject of the next section.

2.7 Voigt profile fitting

Identified absorption lines in the QSO spectra are fitted with Voigt profiles that are first convolved with the COS LSF using the χ^2 minimisation code VPFIT (version 10.0).[15] Convolution with the COS LSF is handled by VPFIT internally, so long as the user provides the tabulated information in an ASCII file. To obtain the appropriate LSF at each wavelength, we interpolate between the tabulated LSFs that are provided by STScI in 50 Å intervals.[16] The COS NUV LSF is described in Ghavamian et al. (2009), and the best FUV LSF characterisation is detailed in Kriss (2011).

For each absorption system identified by the algorithm in Section 2.6, we define fitting regions, then pass both these, and the guessed Voigt parameters (z, $\log N$, b) as input to VPFIT, along with the appropriate LSFs, and run the program.

[14] This condition also roughly limits these identifications to lines that are resolved, or nearly resolved at the COS FUV 'M' grating resolution of ~ 17 km s^{-1}, thus partly eliminating the possibility of identifying lines that are instrumental defects.

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

[16] http://www.stsci.edu/hst/cos/performance/spectral_resolution

Groups of lines at a particular redshift within some arbitrary window (typically a few 100 km s^{-1}) are fit together, but in general we do not tie parameters between different ions[17] (although see Chapter 5). In this way we avoid any assumptions about which groups of ions trace particular absorbing regions, although many will indeed trace the same structures.[18] In some instances, there may be blends of ions from different redshifts that produce compound absorption features. To ensure accurate Voigt profile decompositions, we make every effort to fit these blended absorption lines simultaneously, along with constraints from counterpart, non-blended transitions where available. We also make sure to place fitting regions over areas of the spectrum where absorption is inferred but not detected, as this helps to provide robust upper limits on the best fit column densities.

Each time VPFIT is run, we perform a visual check of the results, along with an inspection of the reduced χ^2 values. We then decide upon whether more or fewer velocity components may be required to give the best fit for any given ion. After some number of trials, we adopt the fit that has the minimum number of velocity components for each ion required to minimise the reduced χ^2 value. Due to the finite velocity resolution of COS, there may be instances where the true number of velocity components in any given absorption complex is higher than the number fitted. Nevertheless, our approach ensures that we make the best use of the information available in the data.

2.8 The absorber sample

After performing a full Voigt profile decomposition of each spectrum, we are left with a catalogue of absorbers for which we list the ions we have identified, together with their best-fit column densities, Doppler broadening parameters, and redshifts. We also assign a reliability flag to each measurement, following a simi-

[17]Note that Voigt parameters between different transitions of the same ion *are* tied in the fitting process, and that in our experience, the wavelength calibration of COS is generally sufficient to allow for this.

[18]This preserves our definition of an 'absorber' to mean a single velocity component of a single ion.

lar scheme to that outlined in Tejos et al. (2014):

- *Secure* ('a'): systems that are detected on the basis of at least two transitions (in the same ion), with $\log N/\Delta(\log N) > 30$, and each transition having an equivalent width significant at the $> 3\sigma$ level.
- *Probable* ('b'): H I systems detected on the basis of Lyα only (after ruling out all other possibilities and with equivalent widths significant at the $> 4\sigma$ level), or metal-line systems detected on the basis of one transition with equivalent widths significant at the $> 3\sigma$ level and with one or more accompanying Lyman series transitions. Both possibilities also with the requirement $\log N/\Delta(\log N) > 30$.
- *Uncertain* ('c'): systems for which $\log N/\Delta(\log N) < 30$ and/or equivalent widths detected at the $< 3\sigma$ level.

Systems labelled 'c' are excluded from the analysis in this thesis. This scheme is also applied to the measurements presented in Danforth et al. (2014). The scheme ensures that we only include absorbers in our sample that are both well constrained and statistically significant. The requirement that H I absorbers detected on the basis of Lyα only must have equivalent widths significant at the $> 4\sigma$ level is motivated in (Danforth et al., 2014), and reduces the number of spurious detections to ~ 3. For the QSOs we have analysed, we present a description and a snapshot of the absorber catalogue in Appendix B, Section B.1 (for the others, see Danforth et al., 2014). The full catalogue will be available for download in a future paper, or can be obtained upon request from the author.

For the purposes of this thesis, excluding Chapter 3, we consider just the O VI samples. We focus on this ion for three main reasons: (i) metal ions in the IGM are closely connected with galaxies, and therefore provide powerful constraints on the interplay between the IGM and galaxies; (ii) O VI is the most commonly observed metal ion in the low-redshift IGM; and (iii) O VI likely traces a wide range of environments and gas phases. The O VI doublet is common because the component transitions have high oscillator strengths, and possess rest-frame wavelengths that make them accessible in the redshift range $0.1 \lesssim z \lesssim 0.7$ with

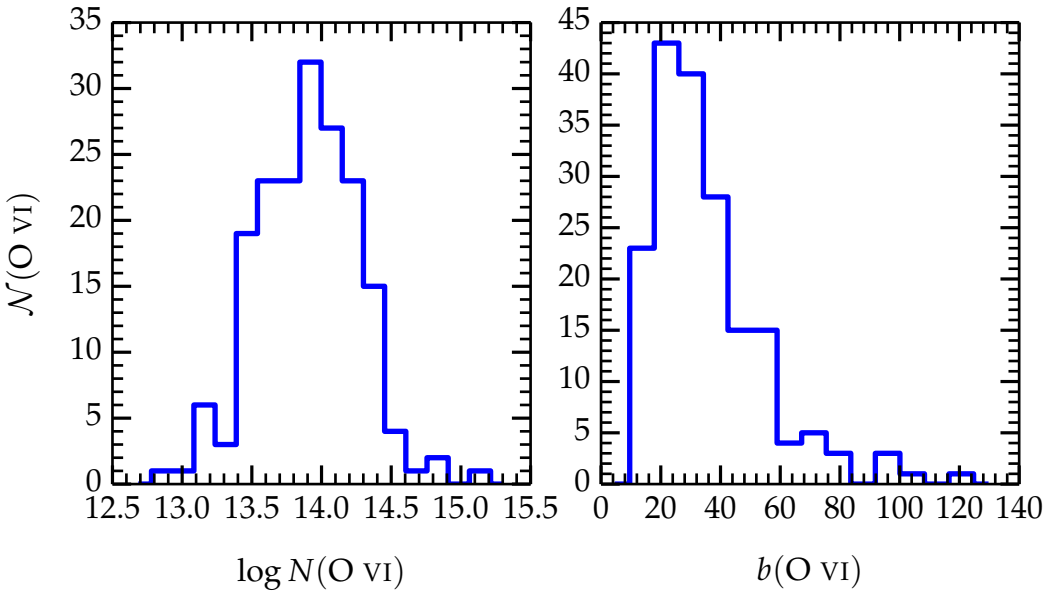

Figure 2.8: Statistics of O VI absorbers in our survey. The left panel shows the histogram of column densities, and the right shows the histogram of Doppler broadening parameters.

current FUV instrumentation. O VI absorbers are thus a convenient tracer of the metal-enriched gas in the IGM. They are thought to trace both cool, photoionized plasmas in the temperature range $10^4 \lesssim T \lesssim 10^5$ K, and hotter, collisionally ionized gas phases at temperatures $10^5 \lesssim T \lesssim 10^7$ K (e.g. Savage et al., 2014; Stocke et al., 2014), the latter of which is commonly referred to as the warm-hot intergalactic medium (WHIM) (e.g. Cen & Ostriker, 1999; Davé et al., 2001; Fukugita & Peebles, 2004), believed to harbour 30 – 40% of all the baryons in the low-redshift Universe. They may also form in more complicated scenarios, e.g. in conductive or turbulent interfaces between gaseous components at multiple temperatures (e.g. Borkowski et al., 1990; Kwak & Shelton, 2010). Therefore, although we suspect that O VI is a good tracer of metal-enriched gas across a large range in gas temperature, we need to bear in mind the many complex formation scenarios for this ion in the interpretation of our results later.

In Figure 2.8, we show histograms of column density and Doppler broadening parameter for our O VI sample. There are a total of 181 O VI systems that possess reliability flags 'a' or 'b'. These absorption systems range over a factor of 100 in column density down to our detection limit ($\log N$(O VI) $\approx 10^{13}$ cm^{-2}), in marked

contrast to H I absorbers that are observed to span $\sim$ 10 orders of magnitude in column density. The number of O VI absorbers drops off fairly rapidly past 10^{14} cm^{-2}, at which point we are typically 100% complete. Doppler broadening parameters show a long tail to high values, and a sharp cut off at $\sim$ 10 km s^{-1}, which roughly corresponds to the spectral resolution of COS. There may be a population of very narrow O VI absorbers, but we are not sensitive to them here.

Chapter 3

A kpc scale outflow associated to a QSO at $z \sim 1$

3.1 Overview

In this chapter, we present Hubble Space Telescope (*HST*)/Cosmic Origins Spectrograph (COS) observations of highly ionized absorption lines associated with a radio-loud quasar (QSO) at $z = 1.1319$. The absorption system has multiple velocity components, with an overall width of ≈ 600 km s^{-1}, tracing gas that is largely outflowing from the QSO at velocities of a few 100 km s^{-1}. There is an unprecedented range in ionization, with detections of H I, N III, N IV, N V, O IV, O IV*, O V, O VI, Ne VIII, Mg X, S V and Ar VIII. We estimate the total hydrogen number density from the column density ratio N(O IV*)/N(O IV) to be $\log(n_{\rm H}/{\rm cm}^{-3}) \sim 3$. Combined with constraints on the ionization parameter in the O IV bearing gas from photoionization equilibrium models, we derive a distance to the absorbing complex of $2.3 \lesssim R \lesssim 6.0$ kpc from the centre of the QSO. A range in ionization parameter, covering ~ 2 orders of magnitude, suggests absorption path lengths in the range $10^{-4.5} \lesssim l_{\parallel} \lesssim 1$ pc. In addition, the absorbing gas only partially covers the background emission from the QSO continuum, which suggests clouds with transverse sizes $l_{\perp} \lesssim 10^{-2.5}$ pc. Widely differing absorption path lengths, combined with covering fractions less than unity across all ions pose a challenge to models involving simple cloud geometries in associated absorption systems. These issues may be mitigated by the presence of non-equilibrium effects, which can be important in small, dynamically unstable clouds. The dynamics and expected lifetimes of the gas clouds suggest that they do not originate from close to

the active galactic nucleus, but are instead formed close to their observed location. Their inferred distance, outflow velocities, and gas densities are broadly consistent with scenarios involving gas entrainment or condensations in winds driven by either supernovae, or the supermassive black hole accretion disc. In the case of the latter, the present data most likely does not trace the bulk of the outflow by mass, which could instead manifest itself as an accompanying warm absorber, detectable in X-rays.

3.2 Introduction

Associated absorption lines (AALs) seen in QSO spectra offer a unique physical perspective on the gaseous environments in the vicinity of QSOs. Many AALs are thought to arise in material that has been ejected from a region close to the super massive black hole (SMBH) (within a few pc, e.g. Nestor et al., 2008; Wild et al., 2008; Muzahid et al., 2013). The resulting outflows might play a major role in the quenching of star formation and in regulating the growth of SMBHs (Silk & Rees, 1998; King, 2003; Scannapieco & Oh, 2004; Di Matteo et al., 2005; Ostriker et al., 2010; Hopkins & Elvis, 2010). Some may arise from material ejected in supernova explosions (e.g. Veilleux et al., 2005). In addition, some AALs appear to probe gas that is part of the host galaxy halo (e.g. Williams et al., 1975; Sargent et al., 1982; Morris et al., 1986; Tripp et al., 1996; Hamann et al., 2001; D'Odorico et al., 2004). In some cases, this gas may eventually condense in the disc to form new generations of stars via Galactic fountain processes (Bregman, 1980; Fraternali et al., 2013). The balance of gas accretion and outflow shapes the galaxy luminosity function and drives the evolution of galaxies (e.g. Benson et al., 2003; Bower et al., 2006). Observations of AALs therefore provide a detailed snapshot of these forces at work. Constraints on the metallicity of these absorbers also provides a direct measure of the star formation and chemical enrichment histories in the centres of active galaxies (Hamann & Ferland, 1993, 1999).

AALs are loosely defined as having absorption redshifts within a few thousand km s^{-1} of the QSO emission redshift, and velocity widths of less than a

few hundred km s^{-1}. These are narrow when compared to the so-called broad absorption lines (BALs), which have velocity widths and displacements from the QSO redshift that often exceed 10^4 km s^{-1} (Weymann et al., 1979; Foltz et al., 1986; Weymann et al., 1991; Trump et al., 2006). The origin of the BALs is presumably in a wind, driven by accretion processes close to the SMBH. However, the exact origin of the AALs is far less clear. In addition, not all AALs are necessarily intrinsic to the QSO (e.g. Tripp et al., 1996; Ganguly et al., 2013). Those that may be intrinsic show: (i) Absorption strength that is seen to vary on time-scales of around a year (e.g. Hamann et al., 1995; Srianand & Petitjean, 2001; Hall et al., 2011; Vivek et al., 2012); (ii) Metallicities $\gtrsim Z_{\odot}$ (e.g. Petitjean et al., 1994; Hamann et al., 1997; Muzahid et al., 2013); (iii) Partial coverage of the QSO accretion disc continuum and/or broad-line region (BLR) (e.g. Barlow & Sargent, 1997; Srianand & Shankaranarayanan, 1999; Gabel et al., 2006; Arav et al., 2008); (iv) The presence of excited fine structure lines (e.g. Morris et al., 1986; Srianand & Petitjean, 2000; Hamann et al., 2001; Edmonds et al., 2011). These properties are rarely seen in intervening absorption-line systems (although see Balashev et al., 2011, for an exceptional case), and so AALs with these properties are believed to trace gas that originates near the QSO, or in the halo of the host galaxy.

AALs have been observed in optical, ultraviolet (UV) and X-ray spectra of local active galactic nuclei (AGN) and QSOs, with the X-ray observations often revealing a plethora of absorption lines and K-shell absorption edges from species with ionization potentials of a few hundred eV (e.g. O VII and O VIII). Collectively, these are usually referred to as 'warm absorbers' (in $\sim$ 50% of Seyfert galaxies; Crenshaw et al., 2003). Many authors have suggested that the presence of warm absorbers is correlated with the detection of AALs and BALs in optical and UV spectra, usually through species like C III, C IV and N V, with ionization potentials $\lesssim$ 100 eV (e.g. Mathur et al., 1994, 1998; Brandt et al., 2000; Kaspi et al., 2002; Arav et al., 2007; Di Gesu et al., 2013). However, at present, it is not clear whether these correlations imply a physical connection between the gas clouds traced by these ions (see for example Srianand & Petitjean, 2000; Hamann et al., 2000, 2013).

To better understand the nature of associated absorbing clouds, more obser-

vations of the most highly ionized UV species (ionization potentials > 100 eV) are required, so that the ionization structure of the absorbing gas can be more extensively characterised. At low redshift, observations of many UV ions are impossible due to the presence of Galactic Lyman limit absorption (the relevant transitions have rest-frame wavelengths < 912 Å). Observations in the optical, which are limited to high redshifts, are complicated by contamination from the Lyman alpha forest, together with a higher incidence rate of Lyman limit systems (e.g. Fumagalli et al., 2013). At intermediate redshifts $0.5 \lesssim z \lesssim 1.5$, the problem of Galactic absorption is virtually eliminated, and the Lyman alpha forest contamination is less severe, making this a profitable redshift range to study highly ionized AALs. Observations must be conducted in the far ultraviolet (FUV), and with the advent of the COS on-board the *HST*, hundreds of QSOs are now observable in this wavelength regime, thanks largely to a sensitivity more than ten times that of the previous generation medium resolution UV spectrograph (Green et al., 2012, Chapter 2). Together with the near ultraviolet (NUV) modes of COS, AALs with ionization parameters of a few, to a few hundred eV are accessible. Detailed diagnostics on the ionization structure of associated gas clouds are thus available in a large number of QSOs for the first time. In addition, coverage of strong transitions due to fine-structure excited states in ions such as O IV and O V (see figure 1 in Arav et al., 2013, for a full summary) provide powerful density diagnostics in highly ionized associated gas clouds, which provide crucial constraints on the physical conditions in and around the absorbing regions.

In this chapter, we present observations of the radio-loud QSO FBQS J0209-0438 obtained with COS. This QSO was targeted as part of a larger programme of observations to study two-point correlation statistics between intergalactic medium (IGM) absorbers and galaxies at $z \lesssim 1$ (PID 12264, PI: S.L. Morris, Tejos et al., 2014, Chapter 6), and it is just one of the QSOs from the sample of 60 that are considered as part of this thesis (see Chapter 2, Table 2.2). A highly ionized system of AALs is present, with complex velocity structure and an overall velocity width ≈ 600 km s^{-1}. We also report the detection of absorption due to the fine-structure O IV* transition. A summary of the observations and data reduction, together

with a characterisation of the rest-frame QSO spectral energy distribution (SED) is presented in Section 3.3. A complete analysis of the AALs; their covering fractions, column densities and line widths is presented in Section 3.4. In Section 3.5 we present the results of photoionization and collisional ionization models, in an attempt to characterise the physical properties of the gas. In particular we examine the ionization state, metallicity, and density of the gas, and use these properties to put constraints on the absorbing geometry and distance from the QSO. In Section 3.6 we present a discussion of these results and draw conclusions.

3.3 Observations of FBQS J0209-0438

The QSO FBQS J0209-0438 (hereafter Q0209) was observed with *HST*/COS in December 2010. The observations made use of both the medium-resolution ($R \sim 18\,000$) FUV and low-resolution ($R \sim 3000$) NUV modes of COS, giving wavelength coverage free from second-order light in the range 1240 – 3200 Å. Four central wavelength settings were used in the FUV, and three in the NUV, to ensure that the resulting spectrum had no gaps. For each central wavelength setting, multiple exposures were obtained at a number of positions offset along the dispersion direction from the nominal one (FP-POS=3; see Table 3.1). Each offset position is separated by ∼ 250 pixels in the FUV channels, and ∼ 52 pixels in the NUV channels, with FP-POS=1, 2 offset to lower wavelengths from FP-POS=3, and FP-POS=4 offset to higher wavelengths. Merging these offsets minimises the effects of fixed-pattern noise in the COS FUV and NUV detectors by effectively dithering around these features, allowing them to be de-weighted (or subtracted) in the final co-added spectra. They are particularly crucial for the FUV modes, which suffer from additional fixed-pattern noise attributable to grid wires that produce shadows on the face of the detector. For more details, including a complete description of the design and in-flight performance of COS, see Chapter 2, Section 2.2, Osterman et al. (2011) and Green et al. (2012).

On the 7th October 2013, under clear skies and excellent seeing conditions (∼ 0.55″), we obtained a near-infrared spectrum of Q0209 with FIRE (Simcoe et al.,

2013), mounted on the Magellan Walter Baade 6.5m telescope at Las Campanas Observatory. In echelle mode, and for the adopted 0.6″ slit, FIRE delivers a continuous spectrum across the wavelength range $0.82 - 2.51\mu$m at a spectral resolution of ~ 50 km s^{-1}. Data were collected in 2×729 s exposures while Q0209 was at an airmass of 1.1. To correct for spectral features arising from the Earth's atmosphere, we also acquired 2 spectra of the A0V star HD25266 at similar airmass, with exposure times of 729 s each.

3.3.1 Data reduction

Individual exposures from COS were downloaded from the Space Telescope Science Institute (STScI) archive and reduced using CALCOS v2.18.5. The boxcar extraction implemented in CALCOS was optimised by narrowing the source extraction box to match the apparent size of the source in the cross-dispersion direction upon inspection of each flat-fielded image. This amounted to 25 pixels for all G130M exposures, and 20 pixels for all G160M and G230L exposures. The background extraction boxes were also enlarged to encompass as much of the background signal as possible, whilst avoiding regions close to the detector edges and the bright object aperture (BOA). We then performed a custom co-addition procedure that works from the `x1d` (`*_x1d.fits`) 1D extracted spectra produced by CALCOS (see Chapter 2, Section 2.4 for full details). These files are listed in Table 3.1 along with the various instrument settings and exposure times.

Data from FIRE were reduced with the FIREHOSE pipeline, which optimally extracts 1D spectra and associated errors in each order from flat-fielded 2D spectral images. The pipeline also computes the wavelength calibration using OH sky lines and ThAr arc lamp spectra obtained after each science exposure. Slit tilts in each order are measured and accounted for in the final wavelength solution, which is in vacuum, and includes the heliocentric correction of 6.8 km s^{-1}. Telluric lines are corrected for in the final spectra, and each order is flux calibrated using the SPEXTOOL software package (Cushing et al., 2004). Finally, the 1D spectra are optimally combined, and each order is merged in a single spectrum. The resulting signal-to-noise (order dependent) ranges between 18 and 36 per spectral pixel.

Table 3.1: A summary of the *HST*/COS observations.

grating	$\lambda_{\rm centre}$ (Å)[a]	FP-POS	$t_{\rm exp}$ (s)[b]	x1d rootname
G130M	1291	2	2321	lbj011ucq
G130M	1291	3	2948	lbj011v4q
G130M	1291	4	2948	lbj011vcq
G130M	1318	3	2948	lbj011vjq
G130M	1318	4	2948	lbj011vsq
G160M	1600	1	2276	lbj012vpq
G160M	1600	2	2948	lbj012vvq
G160M	1600	3	2948	lbj012w2q
G160M	1600	3	2948	lbj012waq
G160M	1600	4	2948	lbj012wiq
G160M	1623	1	2276	lbj013k8q
G160M	1623	2	2948	lbj013keq
G160M	1623	3	2948	lbj013kmq
G160M	1623	3	2948	lbj013ktq
G160M	1623	4	2948	lbj013l0q
G230L	2950	3	2373	lbj014vkq
G230L	2950	4	2985	lbj014vuq
G230L	2635	3	2985	lbj014w3q
G230L	3360	3	2985	lbj014wcq
G230L	3360	4	2985	lbj014wlq

[a] Central wavelength setting.
[b] Exposure time.

3.3.2 Redshift measurement and black hole mass

We measure the redshift of Q0209 using the [O III] $\lambda\lambda$5008,4960 doublet emission lines, seen in the FIRE spectrum. The region around the [O III] and Hβ emission lines is shown in Figure 3.1. An asymmetric profile in the [O III] lines is apparent, with an extended blue wing, indicative of outflowing gas. We decompose the spectrum in this region into the contribution from a power-law continuum, blended Fe II emission lines (Gaussian smoothing over the template of Véron-Cetty et al., 2004) and a multiple Gaussian fit to the [O III] and Hβ emission, using a similar method to that described in Jin et al. (2012). The blue curve shows the resulting best-fit emission model. A two component Gaussian fit to each of the [O III] lines effectively removes the outflowing component. From the line centre in the stronger Gaussian component, we then derive a systemic redshift measurement of $z_{\rm QSO} = 1.13194 \pm 0.00001$, corresponding to a statistical velocity uncertainty of $\Delta v \approx 3$ km s^{-1}. The black hole mass is estimated from the full width at half maximum (FWHM) of the broad Hβ component together with the

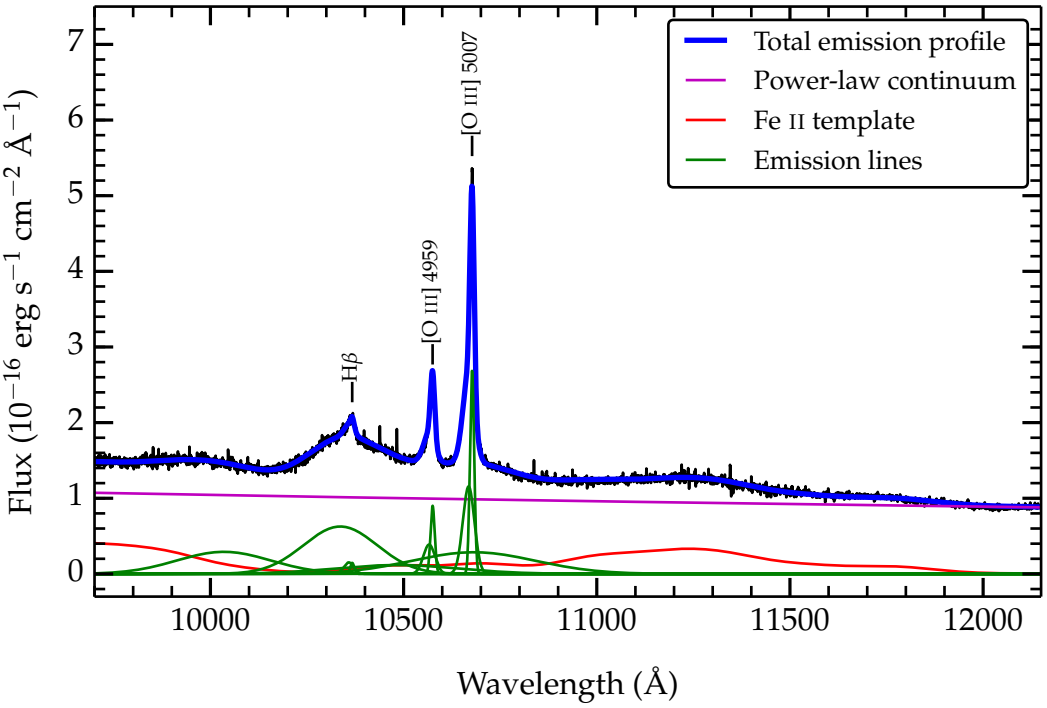

Figure 3.1: A region of the Magellan/FIRE spectrum of Q0209, centred on the [O III] $\lambda\lambda$5008,4960 and Hβ emission lines. The blue curve shows the total fitted emission profile. The model components - a power continuum, Fe II template, and Gaussian emission lines - are shown in magenta, red, and green respectively.

rest-frame 5100 Å flux using equation (3) in Woo & Urry (2002). We find a value $M_{\rm BH} \approx 1.9 \times 10^9 M_\odot$. This value compares favourably with an earlier estimation of $M_{\rm BH} \approx 1.4 \times 10^9 M_\odot$, measured from a Keck/HIRES spectrum using the broad Mg II emission line following the method described in Matsuoka et al. (2013). The latter black hole mass estimate is used as a constraint on the accretion disc models presented in Section 3.3.4, which define the SED of Q0209 used in subsequent photoionization modelling.

3.3.3 The COS spectrum of Q0209

Before performing a complete absorption line analysis on the COS spectrum of Q0209, we first estimate of the unabsorbed QSO pseudo continuum (including emission lines) using the method described in Chapter 2, Section 2.5. The spectrum is then normalised using this continuum. We note that the Lyα emission line

strength is highly uncertain, due to strong absorption disguising the peak in the line. However, the ratio of Lyβ to Lyα line strengths in the fitted spline compares favourably with the same ratio seen in the Shull et al. (2012) *HST*/COS composite spectrum of AGN, and the results presented in this chapter are not sensitive to the exact placement of this peak.

The spline continuum is shown as a blue line on top of the rest frame spectral data in Figure 3.2, corrected for Galactic extinction using the empirical mean extinction curve of Cardelli et al. (1989). We calculate the extinction as a function of wavelength using a Galactic H I column density of 2×10^{20} cm^{-2}, which sits between the measured values of 1.85×10^{20} cm^{-2} (Kalberla et al., 2005) and 2.44×10^{20} cm^{-2} (Dickey, 1990) in this direction. We assume an $E(B-V)$ to $N_{\rm H}$ ratio of 1.5 (Gorenstein, 1975), which gives $E(B-V) = 0.028$. The spectrum is decomposed into the contribution from emission lines, plus that from the accretion disc continuum. We do this by choosing regions of emission-line free continuum, taking the minimum value inferred from the fitted spline in each of these regions, and fitting a power law through the resulting data points, giving a spectral index of $\alpha_\lambda = -0.64$. The flux shortward of $\sim$ 600 Å (1280 Å observed frame) falls to zero due to a Lyman-limit system at $z \simeq 0.39$. A large number of broad emission lines are present, including most of the lines seen in the composite spectrum of Shull et al. (2012) over the same wavelength range. The expected locations of many prominent broad emission lines are labelled.

3.3.4 Spectral energy distribution

For the purposes of photoionization modelling, we construct an SED that extends from the far infra-red ($\sim 10^{-5}$ keV), through to the hard X-ray bandpass ($\sim$ 100 keV), which represents the range in photon energy over which most of the emission is generated by gas accretion. This emission forms the dominant contribution to the ionising photon flux. Data points in the UV are taken from line free regions of the COS spectrum as described in Section 3.3.3. Optical data in the u, g, r, i and z photometric bands come from the SDSS (Ahn et al., 2012),

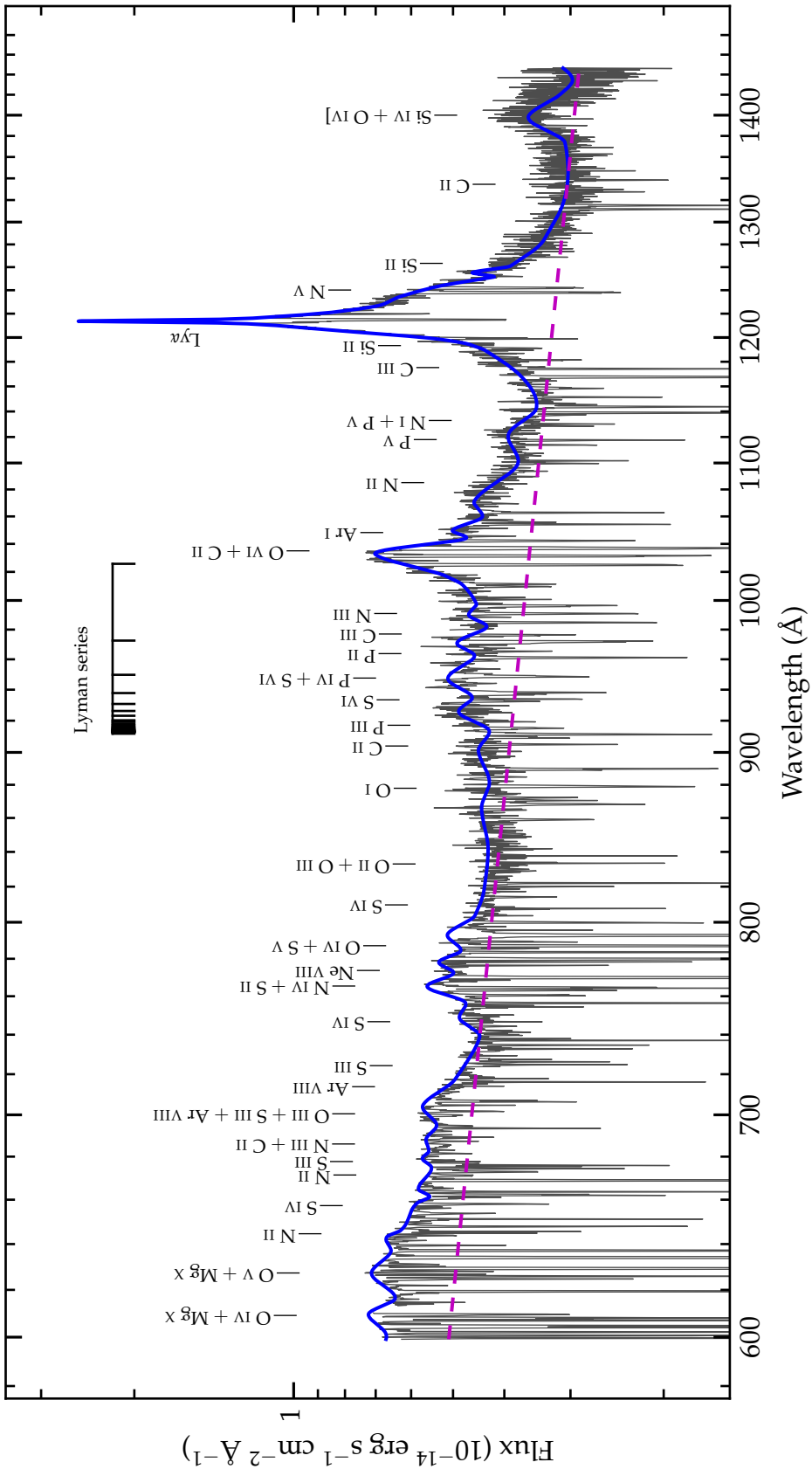

Figure 3.2: The rest-frame *HST*/COS spectrum of Q0209, corrected for Galactic extinction and plotted in 0.5 Å bins. The blue curve shows a spline fit to the total emission profile, as described in Section 2.5. The dashed magenta curve is a power law fit to the continuum profile. A large number of broad emission lines are present. The expected positions of prominent broad emission lines are labelled.

CFHTLS (Cuillandre et al., 2012) and PanSTARRS[1] surveys. Near-infrared data in the J, H, and K_s bands comes from 2MASS (Skrutskie et al., 2006). X-ray points are simulated from the ROSAT all-sky survey (Voges et al., 1999) flux and spectral index using the ROSAT detector response matrix.[2] The data points are corrected for Galactic dust extinction using the method described in Section 3.3.3, and for the Galactic absorption cross-section due to gas, grains, and molecules using the model presented in Wilms et al. (2000). We then fit the data with OPTXAGNF – an energetically self-consistent accretion disc model described fully in Done et al. (2012) – using the spectral fitting package XSPEC.[3] Briefly, the model consists of three main components: (i) a colour-temperature corrected blackbody spectrum powered by the outer regions of the black hole accretion disc; (ii) a soft X-ray excess, attributable to Compton up-scattering of seed photons in the hotter, optically thick inner region of the accretion disc; and (iii) an additional X-ray component formed through Compton up-scattering in a hot, optically thin corona above the disc, creating a power law tail that extends through the hard X-ray bandpass. The model assumes that all the energy used to power these three components is produced through mass accretion. Therefore, the soft and hard X-ray components are physically constrained even though their origin is poorly understood. In modelling the spectrum, we assume a black hole mass of $1.4 \times 10^9 M_\odot$ (see Section 3.3.2).

The resulting rest-frame SED is shown in Figure 3.3. Two models in blue and magenta are shown that fit the data well: one peaking in the soft X-rays, the other in the UV, with reduced χ^2 values of 1.218 and 0.526 respectively. The models differ quite dramatically over the soft X-ray bandpass, representing our ignorance of the true SED shape in this region due to modelling uncertainties on the Galactic extinction across the extreme UV bandpass, and a lack of high quality X-ray observations. Crucially, it is over this energy range where the ions considered in this paper are created (and destroyed). We therefore consider both models in later analysis. The models represent the extremes allowed by the data, and so it is use-

[1] http://www.ps1sc.org
[2] available from http://heasarc.gsfc.nasa.gov/docs/rosat/pspc_matrices.html
[3] http://heasarc.nasa.gov/xanadu/xspec/

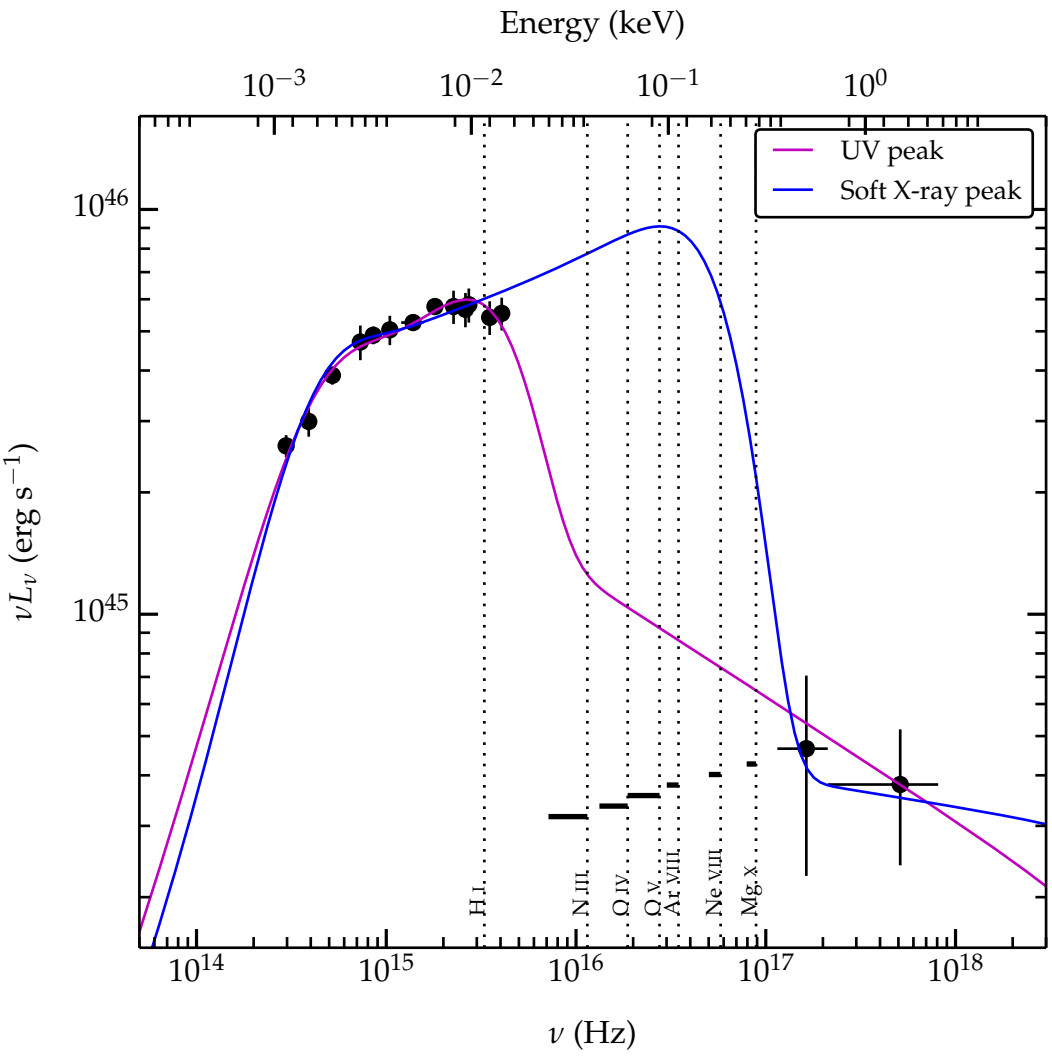

Figure 3.3: Rest-frame model SEDs from OPTXAGNF, fitted to extinction-corrected photometric data and line-free regions of the *HST*/COS spectrum using XSPEC. The magenta curve is the best fitting model, which peaks in the UV. The blue curve is also a good fit to the data, but peaks in the soft X-ray bandpass. Together, these models represent the uncertainty in the SED shape over the soft X-ray region, where there is a large correction for Galactic extinction. Dotted lines represent the ionization destruction potentials for a range of ions that are used as constraints in photoionization modelling. Solid black lines represent the range of energies over which these ions are formed. Crucially, these ions are produced by photons with energies in a region where the SEDs are most markedly different, both in terms of luminosity and spectral slope.

ful to bear in mind that the true SED may be lie somewhere between these two possibilities. Dotted vertical lines represent the ionization destruction potentials for a range of ions later considered in photoionization modelling. Solid black horizontal lines represent the range in energy where these ions are present (extending down to their ionization creation potentials).

3.4 Analysis of the associated absorption

Sections of the COS spectrum of Q0209, together with the spline continuum (blue line) and power-law accretion disc spectrum (dashed magenta line) are shown in Figure 3.4. The spectral resolutions up to, and above, 1750 Å are ~ 16 km s^{-1} and ~ 100 km s^{-1} per resolution element respectively. The plot labels just the most prominent associated absorption troughs, but we report all of the AALs detected with $> 3\sigma$ significance in Table 3.2. Equivalent widths are measured by integrating over the whole absorption trough in each ion (including all discrete velocity components). Some of these troughs are blended with unrelated absorption lines at lower redshifts, making the measured equivalent widths larger than the intrinsic ones. We subtract away the effects of line blending by fitting Voigt profiles with VPFIT[4] (Figure 3.8, Section 3.4.2). The velocity structure across all ions is tied to that of N IV λ765 (Figure 3.5) in the fitting process, based on an empirical (by-eye) similarity between the absorption troughs. This similarity suggests that all the ions we detect in a particular velocity component are formed in regions that are co-spatial. We choose N IV λ765 as a reference, due to a lack of line blending and the fact that almost all components are cleanly resolved in this absorption trough. We have full coverage of the H I Lyman series transitions. Ions searched for, but not detected above a 3σ significance level in the AAL system are C III, O III, O V*, S III, S IV, S VI and Na IX. The expected locations of these particular lines also coincide with unrelated absorption at $z \ll z_{\rm QSO}$ in some cases.

3.4.1 Partial covering

In this section, we look for the presence of partial covering, i.e. indications that the absorbing clouds do not fully cover the background emission from the QSO continuum and/or BLR. The clearest evidence for partial covering comes from flat-bottomed, apparently saturated absorption troughs that do not reach zero intensity (see for example the O V absorption trough in Figure 3.4). If the individual line components are resolved, then these profiles must be caused by optically thick

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

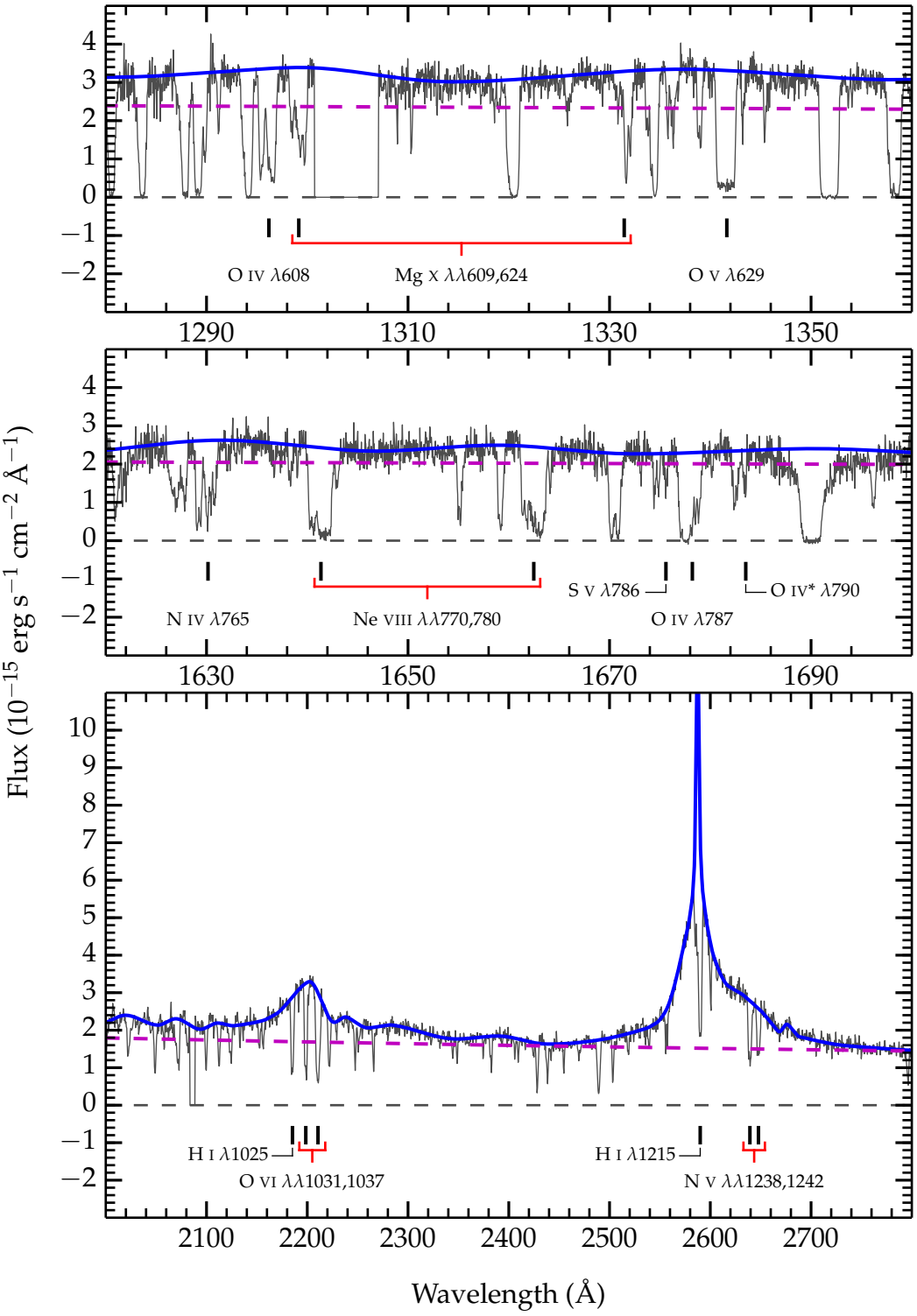

Figure 3.4: The *HST*/COS spectrum of Q0209 in the observed frame, shown in regions containing the most prominent AALs. The blue curve shows the unabsorbed continuum fit including emission lines. The black dashed line shows the zero-flux level. The magenta dashed line shows the power-law accretion disc continuum. Markers indicate the positions of the AALs. Labels indicate the ion (singlet or doublet) and rest-frame transition wavelengths giving rise to those absorption troughs.

Table 3.2: AALs detected at a $> 3\sigma$ significance level, listed first in order of decreasing solar abundance relative to hydrogen, second in order of increasing ionization potential, and third in order of decreasing oscillator strength.

ion	$\lambda_{\rm rest}$ (Å)[a]	IPc (eV)[b]	IPd (eV)[c]	W_λ (Å)[d]
H I	1215.7	–	13.6	2.365 ± 0.031*
	1025.7			1.030 ± 0.032
	972.5			0.967 ± 0.074*
	949.7			0.691 ± 0.045*
	937.8			0.757 ± 0.059*
O IV	787.7			0.937 ± 0.015*
	608.4	54.9	77.4	0.633 ± 0.008*
O IV*	609.8	–	–	0.362 ± 0.009*
	790.2			0.288 ± 0.020**
	790.1			0.288 ± 0.020**
O V	629.7	77.4	113.9	0.914 ± 0.008*
O VI	1031.9	113.9	138.1	1.185 ± 0.040
	1037.6			1.690 ± 0.039*
Ne VIII	770.4	154.2	207.3	1.072 ± 0.011*
	780.3			0.981 ± 0.011*
N III	685.0	29.6	47.4	0.126 ± 0.011**
	685.5			0.092 ± 0.011**
N IV	765.1	47.4	77.5	0.685 ± 0.014
N V	1238.8	77.5	97.9	1.076 ± 0.051
	1242.8			0.768 ± 0.054
Mg X	609.8	328.0	367.5	0.433 ± 0.008*
	625.0			0.362 ± 0.009*
S V	786.4	47.2	72.6	0.185 ± 0.015
Ar VIII	700.2	91.0	124.3	0.165 ± 0.009
	713.8			0.014 ± 0.011

[a] Rest-frame transition wavelength.
[b] Ionization potential for creation.
[c] Ionization potential for destruction.
[d] Observed equivalent width across the entire absorption trough.
* Measured value includes a contribution from blended, unrelated absorption.
** Measured value includes a contribution from absorption lines due to a closely separated transition of the same ion.

absorption *plus* some unabsorbed flux. Under the assumption that the absorbers are spatially homogeneous, the residual flux at an observed wavelength λ in the normalised QSO spectrum may be written as

$$R_\lambda = (1 - C_f) + C_f e^{-\tau_\lambda}, \tag{3.4.1}$$

where τ_λ is the optical depth and C_f the covering fraction, defined as the ratio of occulted to total emitted photons from the background light source(s). Solving

for τ_λ, we have

$$\tau_\lambda = -\ln\left(\frac{R_\lambda - 1 + C_f}{C_f}\right). \tag{3.4.2}$$

For ions with just one transition, estimating the covering fraction is only possible when the line is saturated, in which case the exponential goes to zero, and $C_f = 1 - R_\lambda$. Otherwise, we cannot estimate C_f, as we do not know τ_λ. However, for multiplets, where more than one transition is available, we can eliminate τ_λ by noting that

$$\gamma = \frac{\tau_{\lambda 1}}{\tau_{\lambda 2}} = \frac{f_1 \lambda_1}{f_2 \lambda_2}, \tag{3.4.3}$$

where f_1 and f_2 are the oscillator strengths of each transition. This ratio is close to 2 in the case of doublet lines. For two transitions of the same ion, with residual flux values $R_{\lambda 1}$ and $R_{\lambda 2}$, and covering fractions C_{f1} and C_{f2} respectively, we may then write

$$R_{\lambda 1} = 1 - C_{f1} + C_{f1}\left(\frac{R_{\lambda 2} - 1 + C_{f2}}{C_{f2}}\right)^\gamma \tag{3.4.4}$$

(Petitjean & Srianand, 1999). For simplicity, we can assume that $C_{f1} = C_{f2}$ for each ion, although in general this may not be true (Srianand & Shankaranarayanan, 1999). Complex velocity structure in the broad emission lines (e.g. the presence of both narrow and broad velocity components) can mean that absorbed photons from different parts of a broad line profile will originate from spatially distinct locations (narrow-line region versus broad-line region). Unless the background emission intensity happens to be spatially homogeneous for any given wavelength, this may imply that $C_{f1} \neq C_{f2}$, even for doublets that have a relatively small wavelength separation (see for example the O VI doublet in Figure 3.4, which spans the centre and the blue wing of the Lyβ + O VI emission line). In addition, over the whole wavelength range of the QSO spectrum, there are regions dominated more by the accretion disc continuum than by the BLR, and vice versa. Since the BLR is larger in size than the continuum, if the absorbing clouds have

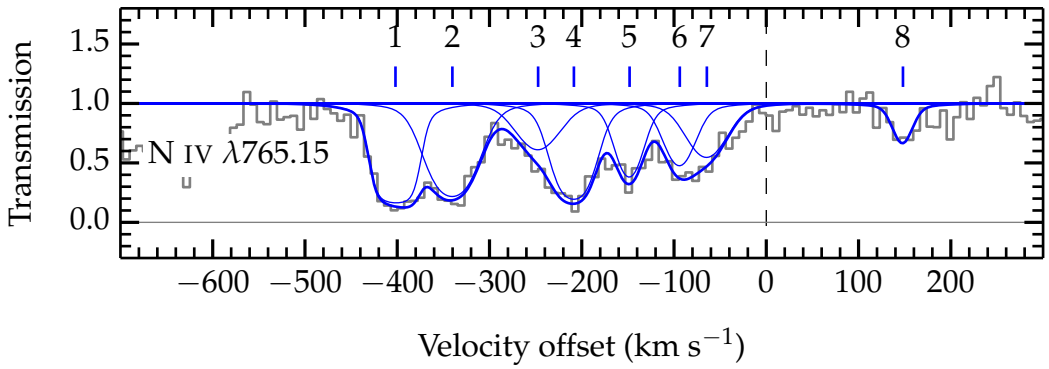

Figure 3.5: Velocity structure in the N IV absorption trough. Voigt components are at −402, −340, −249, −209, −148, −94, −63 and +148 km s^{-1} with respect to the QSO rest frame. The vertical dashed line marks the rest-frame velocity of the QSO. Thin blue lines are individual Voigt profile fits to the data. The thick blue line represents the overall fitted profile.

transverse sizes larger than the continuum region, this can also lead to an inequality. We expect deviations from $C_{f1} = C_{f2}$ to be small in most cases, although the relatively high Lyα emission line flux can lead to situations where the Lyα absorption profile has a smaller apparent optical depth than the Lyβ profile (Petitjean & Srianand, 1999). Nevertheless, we do not see strong evidence for this effect in our data, and we therefore favour a scenario in which there are many clouds with transverse sizes smaller than the continuum size ($\lesssim 4$ light-days; Jiménez-Vicente et al., 2012). In what follows, we assume $C_{f1} = C_{f2}$ within the measurement uncertainties.

Residual flux is clearly present in the saturated absorption troughs of O V λ629 and Ne VIII λ770 (at the blue end), which implies that $C_f < 1$ for these ions (see Figure 3.4). Taking the average normalised residual flux across the flat portions of these profiles gives covering fractions of 0.91 ± 0.01 and 0.93 ± 0.01 respectively (with $\tau \gg 1$ in equation (3.4.1)). In these, and all following covering fraction estimates, the quoted statistical error does not include any contribution from the error on the continuum fit. We also caution that the error bars assume Gaussian statistics, which underestimate the true flux error in absorption troughs where the number of counts in a given bin is low ($\lesssim 100$; Gehrels, 1986). Given the flat profile across the entirety of the O V absorption trough, there is no strong evidence

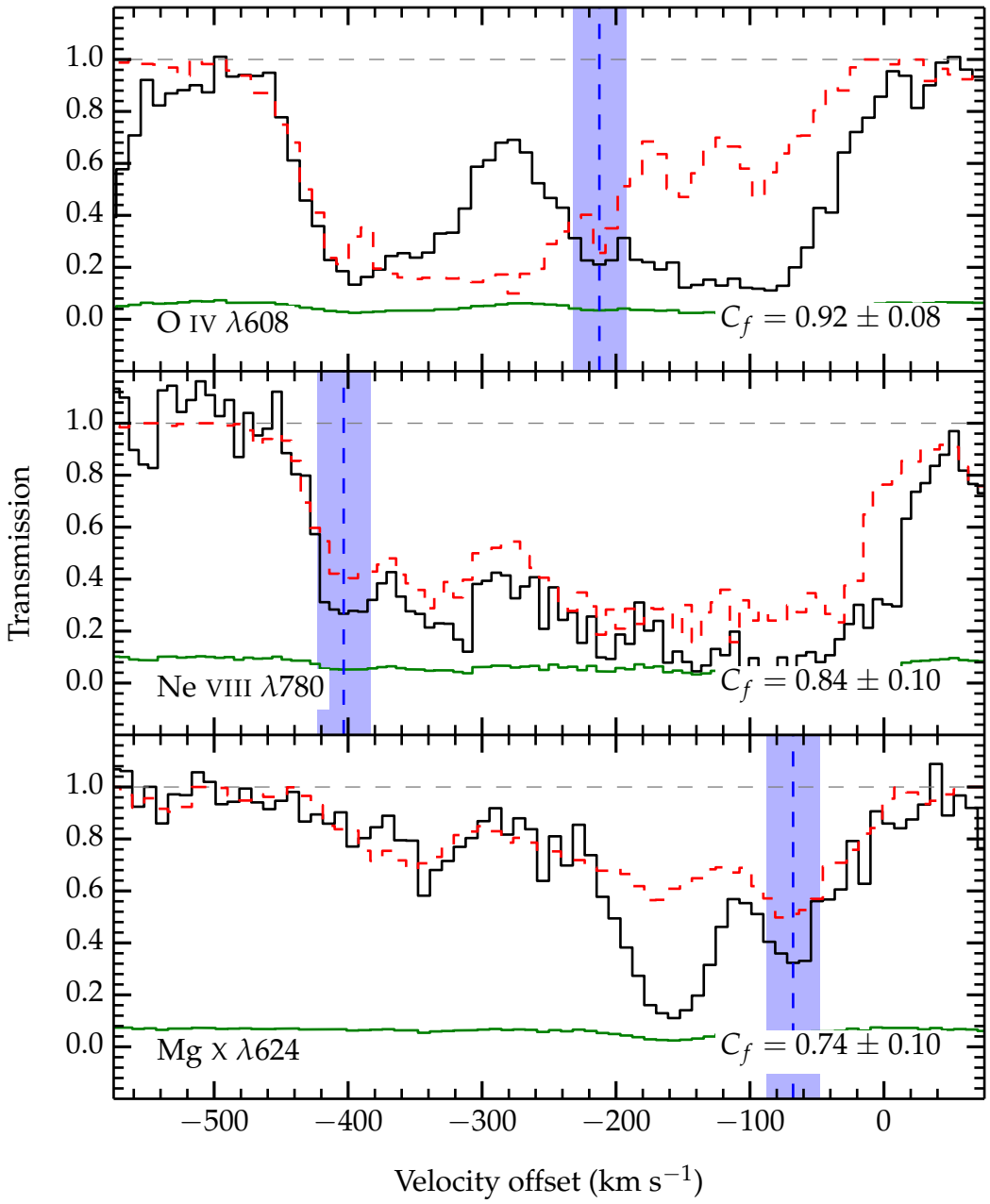

Figure 3.6: Observed (solid black line) compared to predicted (dashed red line) profiles of O IV λ608, Ne VIII λ780 and Mg X λ624 based on apparent optical depth measurements of their stronger counterparts, O IV λ787, Ne VIII λ770 and Mg X λ609 respectively. The deeper observed compared to predicted profiles indicate partial covering of the background emission. This situation is occasionally inverted or accentuated due to line blending (see the text for details). The velocity offset is with respect to the QSO rest frame. The green line is the 1σ error on the transmission in each pixel. Dashed blue lines represent the velocity centroids of the lines not affected (or minimally affected) by blends from unrelated absorption. Vertical shaded regions represent the range of velocities over which covering fractions are calculated (per pixel). Quoted C_f values are the average of those calculated within these regions.

for covering fractions that change across the profiles. We can check this further by examining the apparent doublet ratio for Ne VIII in components that are unsaturated, and unaffected by blending with unrelated absorption lines. There is one component where this measurement is possible in the case of Ne VIII. Covering fractions can also be determined in this way for Mg X and O IV. The former has one velocity component available that meets the aforementioned criteria, and the latter has one that is still mildly affected by line blending. We first calculate the apparent optical depth as a function of velocity across the stronger transition for each ion, then scale these optical depths by γ (equation (3.4.3)) to predict the optical depths in the weaker transitions. The difference between the observed and predicted profiles for the weaker member of each ion can be seen in Figure 3.6. Partial covering results in the predicted profile for the weak member of each ion (red dashed line) having a higher overall transmission than the observed one (black solid line). We note that where there is line blending contaminating the stronger transition, this situation is reversed, and where there is line blending contaminating the weaker transition, this situation is accentuated. The dashed blue lines show the velocity centroids of the components used to measure the covering fractions. We determine covering fractions by numerically solving equation (3.4.4) over a $\sim$ 40 km s^{-1} region centred on each velocity component (blue shaded regions in Figure 3.6), and average the results. This procedure gives covering fractions of 0.92 ± 0.08 for O IV, 0.84 ± 0.10 for Ne VIII, and 0.74 ± 0.10 for Mg X. For O IV, we note that the chosen velocity component is mildly affected by blending, which may add an additional uncertainty on top of the measured one. We also note that the Ne VIII covering fraction determined from this method is consistent with that measured from the saturated blue wing of Ne VIII λ770 within the 1σ uncertainty, and adopt the latter result as the covering fraction for this ion.

For O IV* $\lambda\lambda$609,790 we first note that one of these transitions overlaps with that of Mg X λ609. Therefore, to determine the covering fraction, we first perform a fit to the Mg X λ624 absorption trough, taking into account the covering fraction already determined for this ion, and fixing the velocity structure to that

Table 3.3: Adopted covering fractions.

ion	measurement[a]	C_f[b]	ΔC_f[c]
H I	inferred	0.92	–
N III	inferred	0.92	–
N IV	inferred	0.92	–
N V	inferred	0.91	–
O IV	direct	0.92	0.08
O IV*	inferred	0.92	–
O V	direct	0.91	0.01
O VI	inferred	0.91	–
Ne VIII	direct	0.93	0.01
Mg X	direct	0.74	0.10
S V	inferred	0.91	–
Ar VIII	inferred	0.91	–

[a] Measurements are either direct (based on saturated lines or from comparing line ratios) or inferred (assumed to be the same as that measured directly from an ion with similar ionization potential).
[b] Assuming the covering fraction of the continuum is the same as the covering fraction of the BLR. Requires many clouds smaller than the size of the continuum region.
[c] 1σ statistical uncertainty on the covering fraction (for directly measured values only, not including errors on the continuum fit).

from the fit to N IV λ765 (Figure 3.5). We then re-normalise the spectrum using the calculated Mg X profile, leaving just the absorption signature from O IV*. It is subsequently apparent that the covering fraction for O IV* is consistent with that of O IV. We do not attempt to explicitly calculate the covering fraction in this case, due to the additional uncertainty imposed by subtracting the Mg X absorption. Examining multiplet ratios in N III and Ar VIII reveals that these ions are consistent with a covering fraction of unity, however their detection significance is considerably smaller than for the rest of the ions detected here. Therefore C_f may still be less than 1, as found for ions with better measurements.

For the remaining AALs, covering fractions are even more difficult to determine. In the case of N IV and S V, it is because they are singlet ions (only one transition). In all other cases, it is because the lines fall in the NUV portion of the spectrum, where the lower resolution complicates the process of determining the true residual flux in each line (since individual components are not resolved). The relative contribution of partial covering fractions and resolution effects to the line ratios are very difficult (or impossible) to disentangle. For these lines, the simplest approach is to take another ion with measured covering fraction that is

assumed to trace the same gas, and adopt this covering fraction. We deem the best (available) choice of ion for which this assumption might hold to be that with the most similar ionization potential. We adopt this approach also in the case of N III and Ar VIII. Matching ions under this criterion gives two groups: (O IV, H I, N III, N IV, S V) and (O V, N V, O VI, Ar VIII), where only the first member of each group has a measured covering fraction. Covering fractions for all ions are summarised in Table 3.3. Despite the uncertainties in assigning covering fractions for the absorption lines in our sample, we note that all measured values are high ($C_f \sim 0.9$) with the possible exception of Mg X, and produce statistically good results in Voigt profile fitting (Section 3.4.2). In general, we find that the results of this paper are not sensitive to the precise values of the covering fractions. Indeed, if we take the best measured covering fraction from O V and apply this to all ions, the results we obtain are largely consistent with those obtained later within the measurement uncertainties. In addition, the finding that covering fractions are roughly the same across all ions supports the view that the AALs trace many small gas clouds with transverse sizes less than the continuum size.

3.4.2 Column densities and line widths

To measure the column densities for each ion, we first perform a simultaneous Voigt profile fit to all of the AALs in the medium resolution, FUV part of the spectrum using VPFIT. We do so with the assumption that all ions must share the same 8 separate velocity components, this number having been determined from an independent fit to the N IV λ756 absorption trough (see Section 3.4.1). Voigt profiles are first convolved with the wavelength dependent, non-Gaussian COS line-spread function (LSF) (see Chapter 2, Section 2.7).

For the remaining ions in the low resolution, NUV part of the spectrum, individual components are not resolved, and so these data give very poor constraints on the Doppler broadening (b) parameters, which adds to the uncertainty on the column densities. Crucially, coverage of the H I Lyman series absorption, for which we require well constrained column densities in forthcoming photoionization analysis, is limited to the NUV spectrum. One way of reducing this uncer-

tainty is to find pairs of ions that likely trace the same gas, then require that the b values for these pairs follow some scaling relation, thus reducing the number of degrees of freedom in the χ^2 minimisation. Tying b values sensibly requires knowledge about both the thermal and turbulent motions in the gas. We define b as in Chapter 1, equation (1.4.46). The relationship between b values in two ions, labelled b_1 and b_2, can then be expressed as

$$b_1^2 = \left(\frac{m_2}{m_1}\right) b_2^2 + b_{\rm turb}^2 \left(1 - \frac{m_2}{m_1}\right). \tag{3.4.5}$$

If $b_{\rm turb}$ is zero, then b values can be related by a simple mass scaling. We note that if the gas is photoionized, with a nominal temperature $T \sim 10^4$ K, then initial fits to the FUV data indicate that

$$b_{\rm turb} = \sqrt{b^2 - \frac{2kT}{m}} > 0 \tag{3.4.6}$$

(see Table 3.4). In fact, for $T < 10^5$ K, turbulence dominates the broadening for these lines given the measured b values. Therefore, if the gas is photoionized, we cannot relate b values by the masses of the ions alone.

We can try to determine a plausible range in T by choosing two ions that are assumed to trace the same gas, and whose individual components would then likely possess the same $b_{\rm turb}$. We note that O IV and N IV have almost identical ionization destruction potentials, and very similar ionization creation potentials (54.9 eV and 47.4 eV respectively), making it likely that these ions will satisfy this criterion. We can then require that $b_{\rm turb}$ be the same for each, such that

$$b(\text{N IV}) = \sqrt{b^2(\text{O IV}) + 2kT\left(\frac{1}{m(\text{N})} - \frac{1}{m(\text{O})}\right)}. \tag{3.4.7}$$

We proceed by fitting Voigt profiles to O IV and N IV, requiring that these ions share the same 8 velocity components as before, but additionally force the b values to scale as in equation (3.4.7) for a range of temperatures between $\sim 10^4$ and $\sim 10^6$ K. The resulting reduced χ^2 values on the fit as a function of gas temperature are shown as the black curve in Figure 3.7. Above temperatures of $\sim 10^{5.5}$ K,

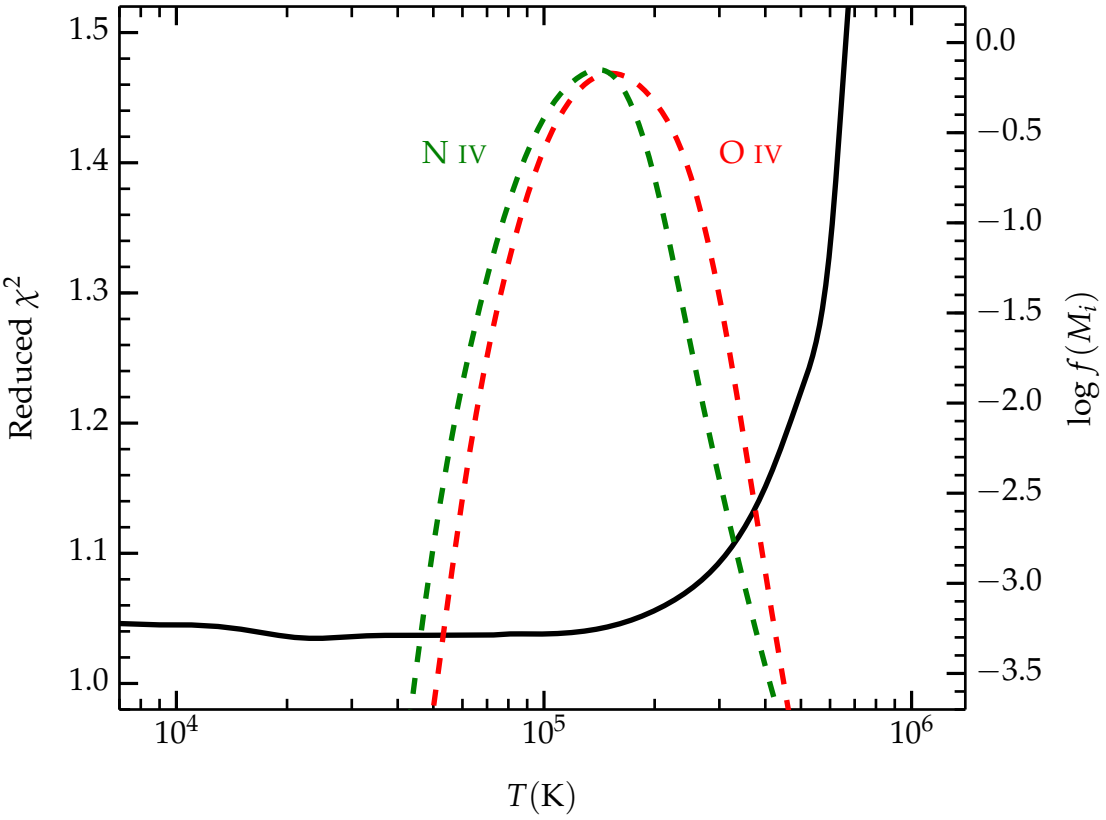

Figure 3.7: Reduced χ^2 values for Voigt profile fits to the N IV and O IV absorption troughs as a function of temperature, assuming their b values scale as in equation (3.4.7). The reduced χ^2 value is minimised in the range $10^4 \lesssim T \lesssim 10^5$ K; temperatures characteristic of photoionized gas. Green and red dotted curves show, for collisional ionization equilibrium, the predicted ion fractions of N IV and O IV respectively as a function of temperature. Both peak at $\sim 10^5$ K, and we are therefore not able to rule out the possibility that the gas is collisionally ionized.

reduced χ^2 values on the fit start to increase dramatically, and we conclude that the data favours gas temperatures less than this value. Turbulence dominates the line broadening at these temperatures, and so scaling b values between pairs of ions based on their masses alone will not be sufficient. Green and red dotted curves in Figure 3.7 show the expected ion fractions of N IV and O IV respectively as a function of temperature for a gas in collisional ionization equilibrium (Mazzotta et al., 1998). Both ion fractions peak at $\sim 10^5$ K, and we are therefore not able to rule out the possibility that the gas is collisionally ionized based on line widths alone.

We proceed by assuming the gas traced by N IV and O IV has a characteristic

temperature of 3×10^4 K (a reasonable value based on Figure 3.7) and use this information to constrain the b values in H I. This procedure minimises the uncertainty in the H I column densities, crucial for later photoionization modelling. We first assume that $b_{\rm turb}$ for N IV be roughly the same as that for O IV and H I, then apply a b scaling between these ions like that in equation (3.4.7). The assumption that N IV and O IV trace the same gas is already well motivated based on the similarity in their ionization potentials. To extend this argument to H I, with ionization potential > 5 times smaller than these ions, we assume that most of the H I is locked up in the same gas as traced by N IV and O IV. We note that this assumption is justified later in Section 3.5 upon consideration of the measured column densities and the ionization fractions derived from photoionization models. It is reassuring to note that the column density measurements resulting from this approach are insensitive to the assumed temperature over the range $10^4 \lesssim T \lesssim 10^5$ K, where the reduced χ^2 values in Figure 3.7 are minimised. Therefore, the gas temperature we assume in the O IV/N IV gas is (nearly) independent of any constraints later obtained in Section 3.5.

All ions are now fitted simultaneously, tying the velocity structure to the 8 components identified in N IV (Figure 3.5) as before, but with an additional b scaling between N IV, O IV and H I as described. All other b values are allowed to float. The resulting fit has a reduced χ^2 value of 1.37 and is shown in Figure 3.8. Transitions are ordered first by atomic mass, and second by oscillator strength. The dashed vertical line indicates the rest-frame velocity of the QSO. Individual blue Voigt profiles represent those components attributable to the labelled ion and transition, whereas red profiles represent components from unrelated blended transitions. Most of these blends are constrained by accompanying transitions of the same ion. The two most prominent blends with O IV λ608, and the blend with Mg X λ625 at $\sim$ 400 km s^{-1}, are assumed to be H I Lyα. The thick blue line represents the overall model absorption trough, and the green points show the residuals on the fit, with the solid black lines marking the $\pm1\sigma$ standard deviation. Line saturation is evident from the Lyα absorption trough, although we note that the column densities are well measured due to a large number of observed

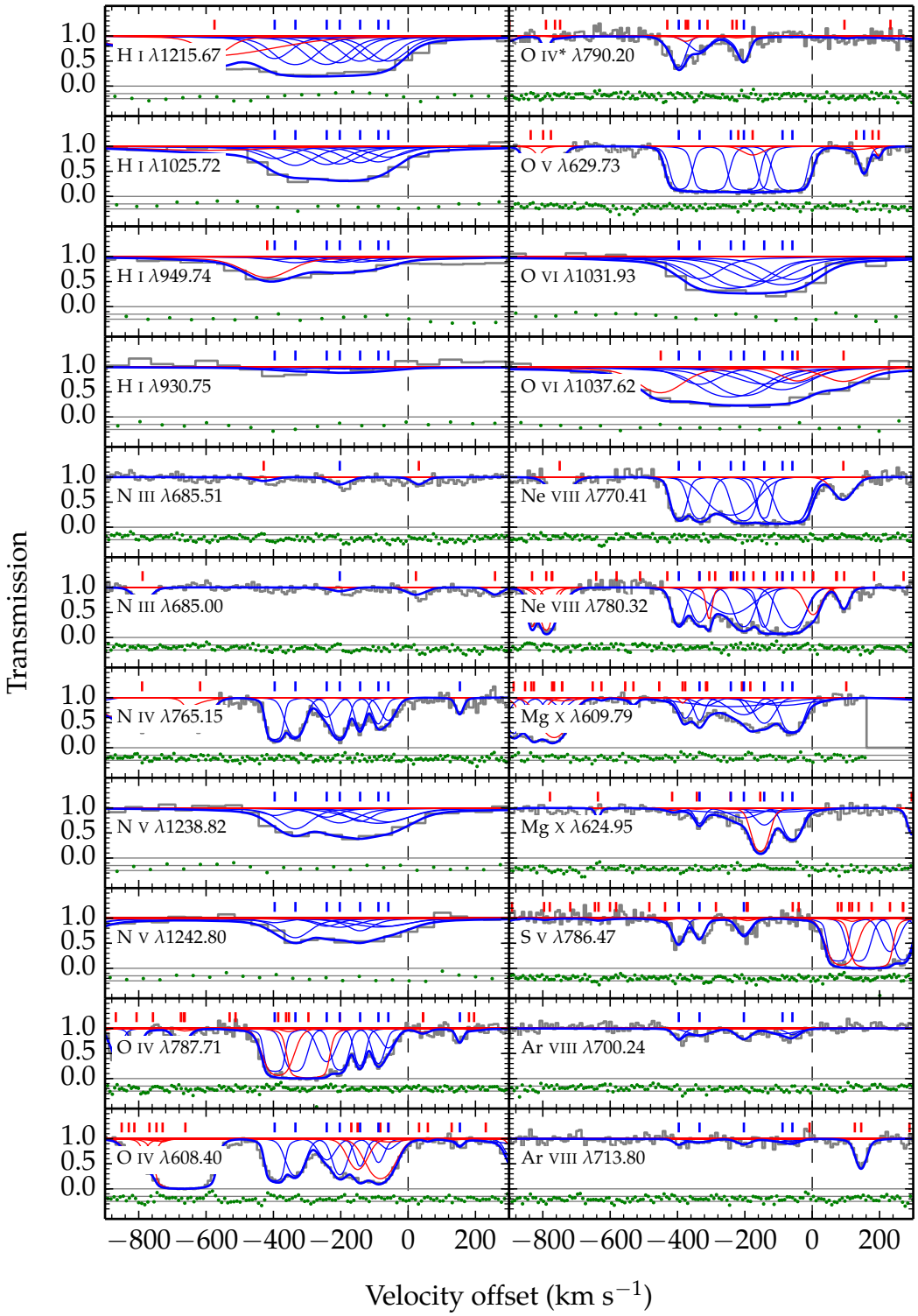

Figure 3.8: Voigt profiles fitted to the AALs in the *HST*/COS spectrum of Q0209. Thin blue lines are the individual velocity components for each labelled ion. Thin red lines are the blended components. The thick blue line in each panel is the total summed profile. The dashed line represents the rest-frame velocity of the QSO. Green points are the residuals between the model and the data, with the solid black lines representing the $\pm 1\sigma$ standard deviation.

Table 3.4: Column density and Doppler broadening parameter measurements of the AALs.

				$\log\left(N/\mathrm{cm}^{-2}\right)$				
ion	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8
H I	14.64 ± 0.32	14.99 ± 0.41	14.92 ± 0.65	14.97 ± 0.97	14.94 ± 0.97	14.83 ± 0.77	14.50 ± 0.29	< 13.46
N III	< 12.99	< 12.99	< 12.99	13.41 ± 0.09	< 12.99	< 12.99	< 12.99	< 12.99
N IV	> 14.11	14.03 ± 0.05	13.42 ± 0.26	14.09 ± 0.11	13.61 ± 0.07	13.50 ± 0.11	13.49 ± 0.10	13.08 ± 0.14
N V	13.83 ± 0.93	> 13.89	> 13.86	> 13.93	> 13.95	> 13.90	> 13.83	< 14.06
O IV	> 14.74	15.02 ± 0.04	14.54 ± 0.22	14.93 ± 0.14	14.59 ± 0.11	14.51 ± 0.09	14.12 ± 0.15	13.68 ± 0.12
O IV*	14.45 ± 0.06	14.10 ± 0.12	< 13.46	14.22 ± 0.07	< 13.46	< 13.46	< 13.46	< 13.46
O V	> 14.24	> 14.34	> 14.39	> 14.24	> 14.14	> 14.22	> 14.22	13.62 ± 0.09
O VI	> 13.90	> 14.21	> 14.22	> 14.25	> 14.34	> 14.35	> 14.28	< 14.12
Ne VIII	14.94 ± 0.07	14.91 ± 0.19	15.06 ± 0.29	15.10 ± 0.15	> 14.83	> 15.01	> 14.91	< 13.48
Mg X	< 13.46	14.50 ± 0.12	14.80 ± 0.28	14.68 ± 0.39	14.78 ± 0.19	14.54 ± 0.54	15.29 ± 0.09	< 13.46
S V	13.14 ± 0.07	12.97 ± 0.08	< 12.40	12.90 ± 0.09	< 12.40	< 12.40	< 12.40	< 12.40
Ar VIII	13.19 ± 0.13	13.28 ± 0.13	< 12.82	13.34 ± 0.09	< 12.82	13.26 ± 0.22	13.04 ± 0.29	< 12.82
				$b\left(\mathrm{km\ s}^{-1}\right)$				
ion	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8
H I	24	34	35	29	26	28	33	–
N III	–	–	–	34 ± 12	–	–	–	–
N IV	–	27 ± 3	27 ± 13	20 ± 3	15 ± 4	18 ± 6	25 ± 6	10 ± 7
N V	18 ± 106	–	–	–	–	–	–	–
O IV	–	26	27	20	15	18	25	10
O IV*	20 ± 4	33 ± 14	–	17 ± 6	–	–	–	–
O V	–	–	–	–	–	–	–	13 ± 5
O VI	–	–	–	–	–	–	–	–
Ne VIII	19 ± 2	26 ± 7	77 ± 76	29 ± 12	–	–	–	–
Mg X	–	14 ± 6	171 ± 78	54 ± 33	31 ± 12	109 ± 96	25 ± 5	–
S V	18 ± 5	17 ± 6	–	23 ± 8	–	–	–	–
Ar VIII	16 ± 9	41 ± 19	–	25 ± 9	–	49 ± 22	22 ± 19	–

transitions in the H I Lyman series.

Column densities and Doppler broadening parameters are listed on a component by component basis ($v_1 - v_8$, ascending in velocity offset), together with their 1σ error bars in Table 3.4. For components where no absorption line is detected in any given ion, we calculate the upper bound on the equivalent width at the 3σ significance level, derived using equations (4) – (5), (7) and (9) – (10) in Keeney et al. (2012), then perform the conversion to column density assuming a linear curve of growth. For column densities with 1σ uncertainties greater than 1 (typically when $\tau \gg 1$, i.e. where the lines are saturated), we quote lower limits on the column densities based on the apparent optical depth at the line centres (calculated using equation (3.4.2)) and assuming a b value of 25 km s^{-1}, which is approximately typical of the well-measured lines. We do not list b values for components that have upper or lower limits on the column densities. Error bars are not presented for inferred b values.

3.5 Properties of the associated absorbers

In the following section, we first present constraints on the electron number density in the associated absorbers based on an analysis of the fine-structure transition O IV*. We then consider both photoionization and collisional ionization equilibrium (CIE) models, and use these to put constraints on the properties of the gas clouds. Models are generated using version c13.00 of Cloudy, described in Ferland et al. (2013). We note that, in general, absorbers may not be in ionization equilibrium (e.g. Oppenheimer & Schaye, 2013a,b). This possibility is discussed later in Section 3.6.2.

3.5.1 Electron number density in the absorbing clouds

In this section we present analysis on the absorption due to the fine structure, metastable transitions in O IV, which enables us to estimate the electron number density in the clouds giving rise to the absorption by O IV. First, we note that O IV* arises from doublet fine-structure ($J = 1/2$ and $3/2$) in the ground state,

which should behave approximately as a two-level atom, where the level populations are controlled by collisional processes and forbidden radiative decays (Bahcall & Wolf, 1968). If we neglect stimulated emission, the only acting processes are collisional excitation, collisional de-excitation and radiative decay. The energy level spacing corresponds to 25.91 μm, or 1.2×10^{13} Hz, so we find this to be a fair approximation on the basis of Figure 3.3, together with the fact that stimulated emission is extremely forbidden. We denote the ground state as level 0, the excited state as level 1, and let n_j be the number density (cm^{-3}) of O IV in level j. If we assume that collisional excitation is dominated by electrons, then the population of the excited state must satisfy

$$\frac{\mathrm{d}n_1}{\mathrm{d}t} = n_e n_0 k_{01} - n_e n_1 k_{10} - n_1 A_{10} \tag{3.5.8}$$

(Draine, 2011), where n_e is the electron number density (cm^{-3}), k_{01} and k_{10} are the upward and downward rate coefficients ($\mathrm{cm}^3\ \mathrm{s}^{-1}$) respectively, and A_{10} is the spontaneous decay rate (s^{-1}). For a steady state ($\mathrm{d}n_1/\mathrm{d}t = 0$), we then require

$$\frac{N(\mathrm{O\ IV^*})}{N(\mathrm{O\ IV})} = \frac{n_e k_{01}}{n_e k_{10} + A_{10}}, \tag{3.5.9}$$

where we have now replaced the number densities by the observed column densities. Note that this implicitly assumes O IV and O IV* trace the same gas. We can write k_{01} in terms of k_{10} as follows:

$$k_{01} = \frac{g_1}{g_0} k_{10} e^{-E_{10}/kT}, \tag{3.5.10}$$

where g_0 and g_1 are the level degeneracies, E_{10} is the energy level difference, and T is the kinetic temperature of the gas. We can additionally define the critical electron density, at which the collisional de-excitation rate equals the radiative de-excitation rate:

$$n_{\mathrm{crit}} \equiv \frac{A_{10}}{k_{10}}. \tag{3.5.11}$$

Table 3.5: Total hydrogen number density in each velocity component.

$\log\left(n_{\rm H}/{\rm cm}^{-3}\right)$							
v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8
$\lesssim 3.54$	2.80 ± 0.13	$\lesssim 2.63$	3.03 ± 0.17	$\lesssim 2.70$	$\lesssim 2.66$	$\lesssim 3.09$	$\lesssim 3.64$

The electron number density can then be written as

$$n_e = n_{\rm crit}\left(\frac{N({\rm O\ IV})}{N({\rm O\ IV^*})}\frac{k_{01}}{k_{10}} - 1\right)^{-1}. \tag{3.5.12}$$

Assuming a temperature of 10^4 K, we can write

$$k_{10} = 8.629 \times 10^{-8}\frac{\Omega_{10}}{g_1}\ {\rm cm}^3\ {\rm s}^{-1}, \tag{3.5.13}$$

where Ω_{10} is the collision strength connecting levels 1 and 0. For electrons at this temperature, the collision strength is $\Omega_{10} = 2.144$ (Tayal, 2006). We take $A_{10} = 5.19 \times 10^{-4}$ s^{-1} from the NIST atomic spectra database.[5] The resulting dependence of n_e on the ratio $N({\rm O\ IV^*})/N({\rm O\ IV})$ is shown in Figure 3.9. Comparing to figure 12 in Arav et al. (2013), we find an excellent agreement. For a highly ionized plasma, assuming solar metallicity and abundances, we can relate n_e to the hydrogen number density $n_{\rm H}$ as

$$n_e = n_{\rm H}\left(1 + 2\frac{n_{\rm He}}{n_{\rm H}} + \sum_{i\geq 3} Z_i\frac{n_{Z_i}}{n_{\rm H}}\right) \approx \frac{1+X}{2X}n_{\rm H}, \tag{3.5.14}$$

where Z_i is the atomic number of element i, and X is the mass fraction in hydrogen. For a value $X = 0.71$, we have $n_e \simeq 1.2 n_{\rm H}$. We calculate $n_{\rm H}$ in the gas traced by O IV for each velocity component using equation (3.5.12), and summarise the results in Table 3.5. Where there is a lower limit on the O IV column density, or an upper limit on the O IV* density, this results in an upper limit on $n_{\rm H}$. We caution that these upper limits are only approximate, as they are sensitive to the adopted significance level used to calculate limiting equivalent widths, and the assumed

[5] http://www.nist.gov/pml/data/asd.cfm

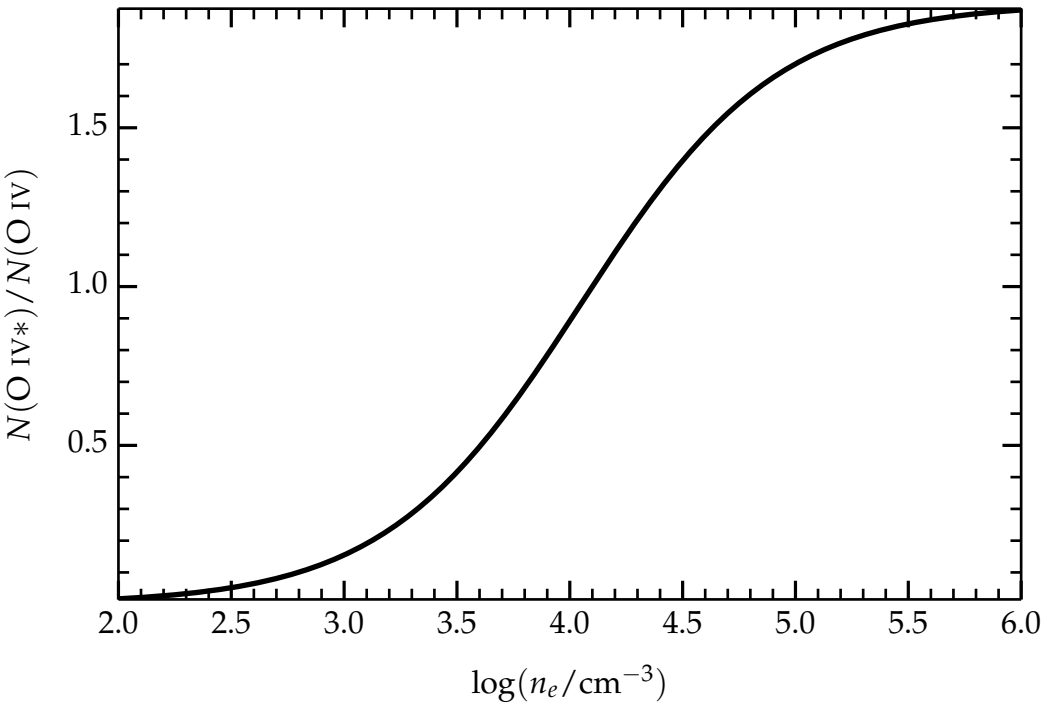

Figure 3.9: The electron number density, n_e, as a function of the column density ratio between O IV* and O IV based on predictions from a theoretical level population assuming a temperature of 10^4 K.

b value used to calculate the lower limit on the O IV column density. We find that a value $\log(n_\mathrm{H}/\mathrm{cm}^{-3}) \sim 3$ is representative for the AAL region as a whole.

3.5.2 Photoionization analysis

In the following, we assume that the gas clouds are in photoionization equilibrium with either of the QSO SEDs determined in Section 3.3.4. The two main parameters that best describe the photoionization structure are the total hydrogen column density (N_H) and the ionization parameter (U). The latter is defined as the dimensionless ratio between the number density of hydrogen atoms, and the number density of photons that ionize hydrogen at the illuminated face of the absorbing gas clouds. We express this as

$$U \equiv \frac{1}{4\pi c R^2 n_\mathrm{H}} \int_{\nu_\mathrm{LL}}^{\infty} \frac{L_\nu}{h\nu} d\nu, \tag{3.5.15}$$

where ν is the frequency, c is the speed of light, L_ν is the luminosity density of the QSO (erg s^{-1} Hz^{-1}), $\nu_{\rm LL}$ is the frequency corresponding to the Lyman limit (912 Å), R is the radial distance between the absorber and the QSO, and $n_{\rm H}$ is the total hydrogen density (i.e. H I + H II). The equations of ionization and thermal balance are then solved using Cloudy. We run the code multiple times to form three-dimensional grids of predicted quantities over (U, $N_{\rm H}$, Z) parameter space, where Z is the metallicity of the gas, normalised to solar metallicity. We vary the parameters U and $N_{\rm H}$ in steps of 0.1 dex, and Z in steps of 1 dex, assuming a constant total hydrogen density. The results from these grids are the subject of this section and Section 3.5.4.

Shown in Figure 3.10 are theoretical ionization fractions of H I and various metal ions for which we have reliable column density estimates as a function of U. We show model curves for clouds in photoionization equilibrium with both the 'UV peak' and 'Soft X-ray peak' SEDs (see Section 3.3.4). These results are not sensitive to the abundances used in the calculations (in this case we have used solar abundances), and neither are they sensitive to the total column densities. This is because the model clouds are optically thin in the Lyman continuum, which means that there are no steep gradients in ionization (e.g. due to shielding from an H II – H I recombination front). We note that this situation is a good approximation to the real one, given the absence of absorption at the Lyman limit. We calculate the column density ratios N(O IV)/N(O V) and N(Ne VIII)/N(Mg X) in the components where those ratios are best measured, then find the range in ionization parameter over which each is predicted within their 1σ uncertainties. For N(Ne VIII)/N(Mg X), we scale by relative solar abundances. The resulting constraints on U are plotted as a series of vertical bars.

From Figure 3.10, it is clear that the gas has a range in ionization parameter that covers around two orders of magnitude. At a fixed distance R from the QSO, this corresponds to a range in gas density that covers around two orders of magnitude. Measuring the same column density ratios in different velocity components to those plotted (including column density lower limits) gives constraints on the ionization parameter that are fully consistent with those obtained above.

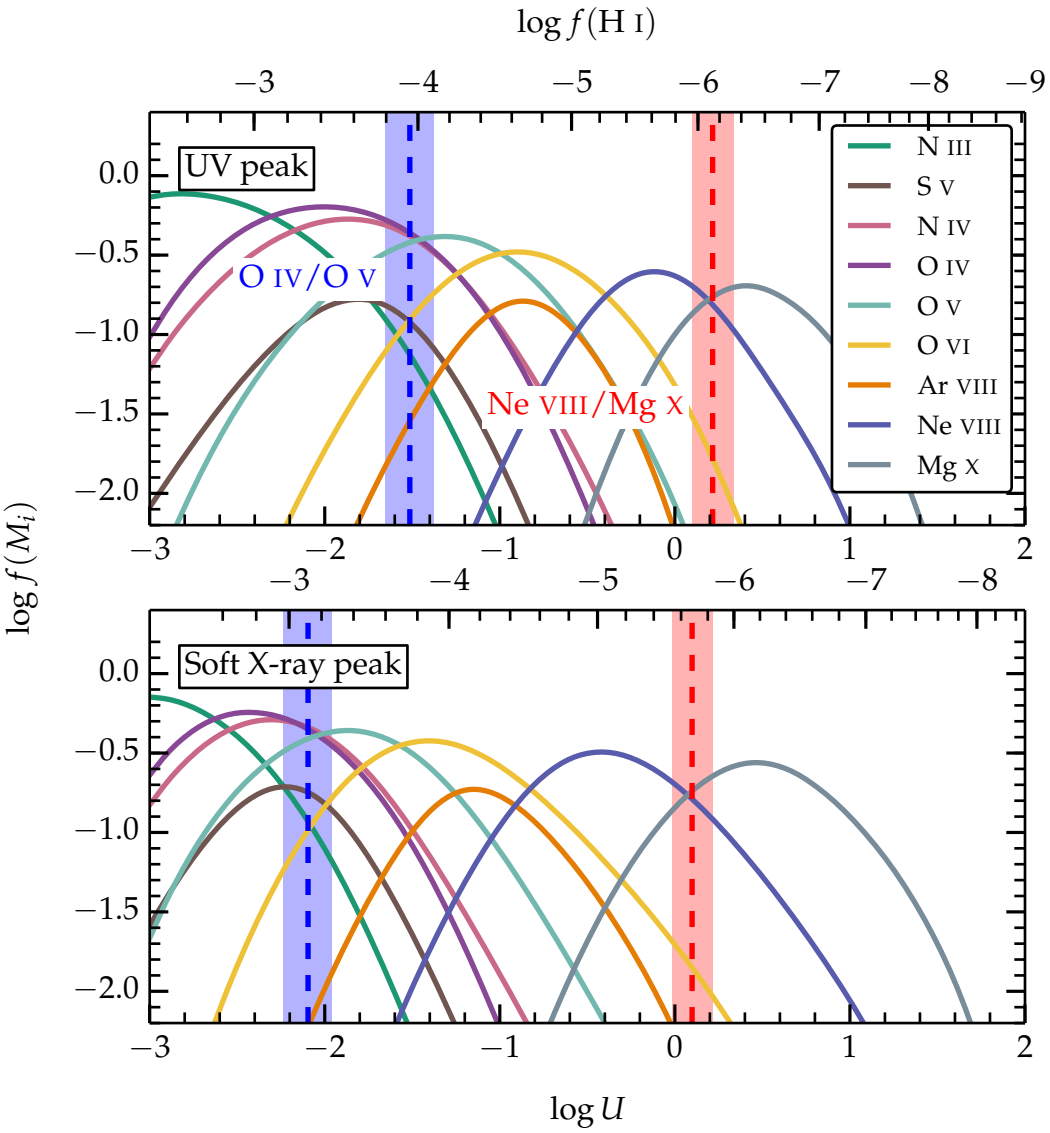

Figure 3.10: Theoretical ionization fractions for a range of metal ions $f(M_i)$ in optically thin clouds in photoionization equilibrium with the 'UV peak' SED (top panel) and the 'Soft X-ray peak' SED (bottom panel), plotted as a function of ionization parameter. Curves are colour coded by ion. The H I fraction is shown on the top axis. Vertical bars indicate the range in ionization parameter allowed for by measured column density ratios in components where those ratios are best measured. For N(Ne VIII)/N(Mg X) we scale by relative solar abundances. These ratios indicate there is a range in ionization parameter covering nearly two orders of magnitude under photoionization equilibrium.

This range ionization parameter is therefore representative for the AAL region as a whole. This range can be considered a lower limit, since we cannot rule out the presence of even more highly ionized species that lie outside the wavelength coverage afforded by COS, which may infer even higher ionization parameters.

3.5.3 Collisional ionization equilibrium

We next consider a situation whereby the AAL clouds are in collisional ionization equilibrium (CIE). To simulate CIE, we run Cloudy multiple times with $\log U = -5$ (so that the effect of the radiative field is negligible), holding temperature constant for values in the range $4.5 \leq T \leq 6.5$. Shown in Figure 3.11 are theoretical ionization fractions of various metal ions for which we have reliable column density estimates under CIE as a function of temperature. Like Figure 3.10, the results are not sensitive to the specific abundances adopted, or the total column density. Vertical bars indicate the range in temperature permitted according to the best-measured column density ratios N(O IV)/N(O V) and N(Ne VIII)/N(Mg X) within their 1σ uncertainties. Again, for N(Ne VIII)/N(Mg X) we scale by relative solar abundances. It is clear that temperatures cover more than an order of magnitude in the AAL region if the gas is in CIE. Similarly to the case of photoionization, constraints on the temperature are consistent between different velocity components.

To determine whether CIE is allowed by the data, we model clouds illuminated by the 'UV peak' incident continuum, assuming an H I column density of 10^{15} cm^{-2} (matching the observed value in velocity component v_2), a total hydrogen density of $\log(n_{\rm H}/{\rm cm}^{-3}) = 2.8$, and solar metallicity, for a range of ionization parameters U, holding T constant, and repeating to create a grid of values (U, T). We then plot the predicted O IV column density as a function of T for a range of values of U in Figure 3.12. CIE is achieved in the limit of high T and/or low U. The horizontal grey bar represents the 1σ constraint on the O IV column density in velocity component v_2. The dashed blue vertical line and accompanying arrow represents the lower limit on the temperature based on the column density ratio O IV/O V. This is then an indication of the characteristic temperature of the gas hosting O IV in CIE. Temperatures to the right of the red dashed line at $10^{5.5}$ K have already been ruled out based on the observed line-widths (see Figure 3.7).

Given the lower and upper bounds on the gas temperature, it is clear that the predicted O IV column density in Figure 3.11 is at least an order of magnitude larger than that observed. All theoretical curves match the observed O IV column

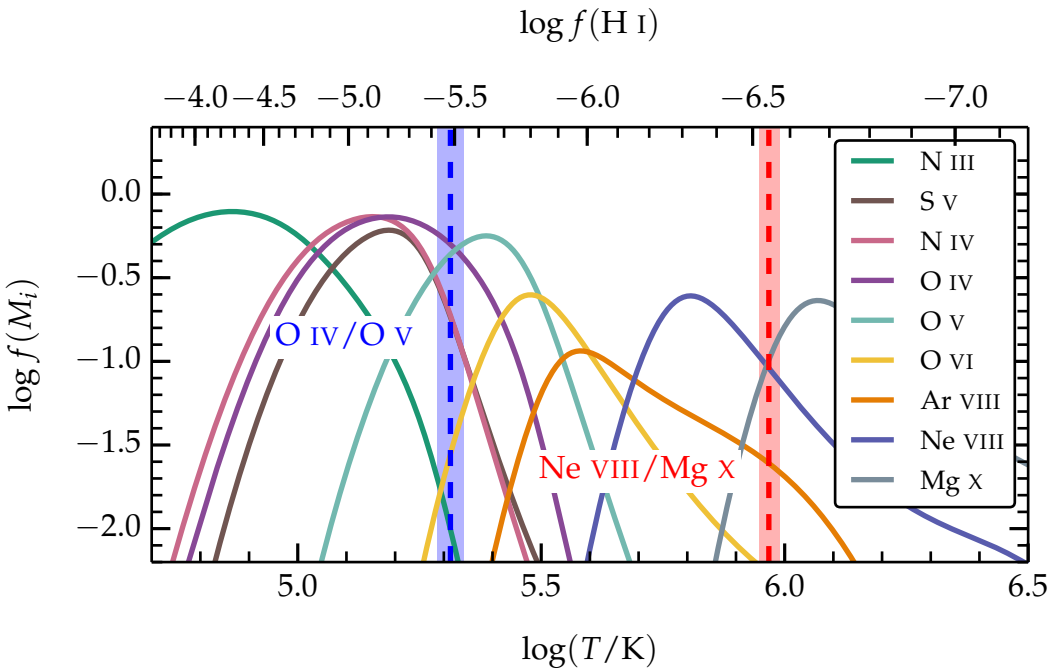

Figure 3.11: Theoretical ionization fractions for a range of metal ions $f(M_i)$ under CIE as a function of temperature. Curves are colour coded by ion. Vertical bars indicate the range in temperature allowed for by measured column density ratios in components where those ratios are best measured. For N(Ne VIII)/N(Mg X), we scale by relative solar abundances. These ratios indicate a range covering more than an order of magnitude in gas temperature under CIE.

density for $T \lesssim 10^{4.85}$ K. We note that decreasing U further has no effect on this result, since by $\log U = -5$, the solution has converged to a value $T \approx 10^{4.85}$ K. The assumed H I column density is close to the maximum allowed by the data, although this value would have to be reduced by at least an order of magnitude to give agreement between the predicted temperatures. Similarly, the metallicity would have to be lowered by at least an order of magnitude. In photoionization equilibrium, the ionization parameter for the gas traced by O IV is $\log U \sim -2$, which upon inspection of Figure 3.12 gives a temperature of $T \approx 10^{4.5}$ K, fully consistent with that predicted by the photoionization models. Changing the incident continuum to the 'Soft X-ray peak' model gives similar results. We therefore conclude that the AAL gas traced by O IV (and ions with similar ionization potential) is predominantly photoionized.

The data also provide indications that the most highly ionized species are predominantly photoionized. From Section 3.4.2, we found that the line widths in

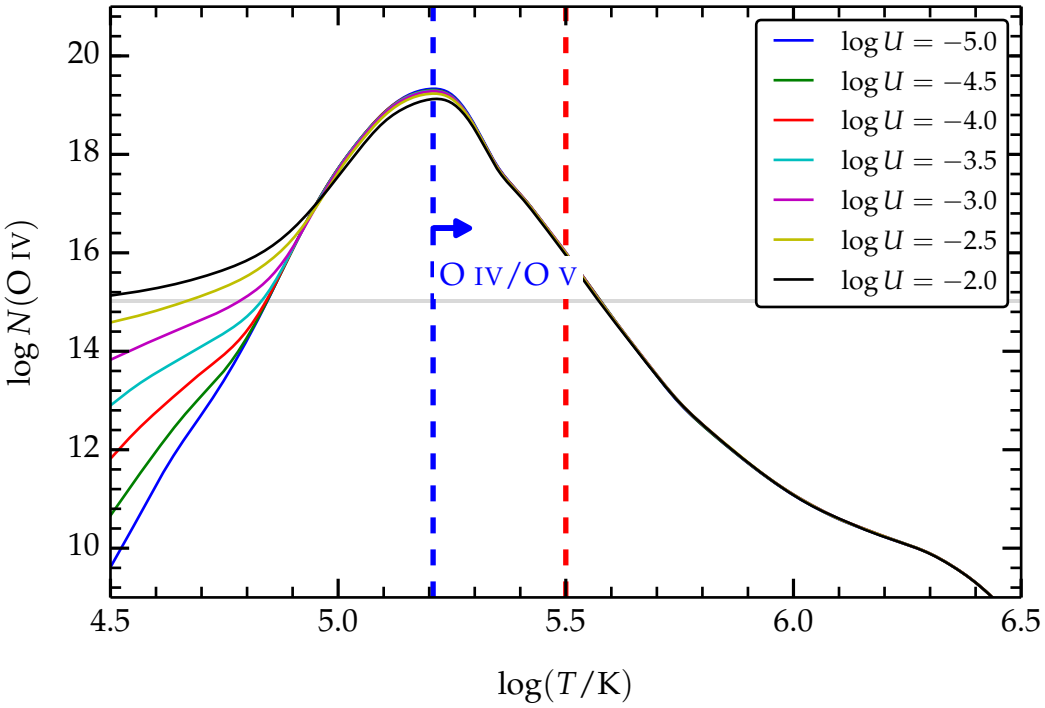

Figure 3.12: The predicted O IV column density as a function of gas temperature for a range of ionization parameters and a fixed H I column density of 10^{15} cm^{-2}. The dashed blue vertical line and accompanying arrow denotes a lower limit on the temperature of the gas clouds giving rise to the majority of the O IV absorption under CIE, inferred from the ratio N(O IV)/N(O V). The vertical red dashed line marks the maximum temperature allowed from an analysis of the absorption line widths. The horizontal grey shaded bar represents the 1σ constraints on the O IV column density from velocity component v_2. Theoretical curves match this column density at temperatures nearly an order of magnitude lower than those inferred from CIE models, favouring a scenario in which the AAL gas traced by O IV is predominantly photoionized.

O IV and N IV are dominated by the effect of turbulence. Taking velocity component v_2 as an example, given the measured line width in N IV, and a maximum temperature allowed by the data of 3×10^5 K, we find that $b_{\rm turb} \gtrsim 19$ km s^{-1}. We'll assume that this turbulent contribution to the line widths in velocity component v_2 is also representative for Ne VIII and Mg X, which is supported by the similar velocity structure across all ions, suggesting that the gas traced by N IV and O IV is co-spatial with that traced by Ne VIII and Mg X. Then, given the predicted temperature $T \approx 10^6$ K for CIE, we should expect $b \gtrsim 35$ km s^{-1} and $b \gtrsim 32$ km s^{-1} for Ne VIII and Mg X respectively. These are in tension with the measured values, which are below these predictions. We see a similar situation across all other

velocity components where we have good constraints on the line widths. Therefore, the data suggest that the AALs are predominantly photoionized overall, and we can largely rule out CIE as an ionization mechanism for the gas traced by the AALs.

3.5.4 Gas metallicity and total column density

The metallicity, $[M/\mathrm{H}]$, of the gas giving rise to the AALs can be expressed as

$$\left[\frac{M}{\mathrm{H}}\right] = \log\left(\frac{N\left(M_i\right)}{N\left(\mathrm{H\ I}\right)}\right) + \log\left(\frac{f\left(\mathrm{H\ I}\right)}{f\left(M_i\right)}\right) + \log\left(\frac{\mathrm{H}}{M}\right)_{\odot} \tag{3.5.16}$$

(Hamann & Ferland, 1999), where $(\mathrm{H}/M)_{\odot}$ is the solar abundance ratio of hydrogen to some metal species M, $N(\mathrm{H\ I})$ and $f(\mathrm{H\ I})$ respectively are the column density and ionization fraction in H I, and $N(M_i)$ and $f(M_i)$ respectively are the column density and ionization fraction in some ion M_i of metal species M. If the gas is well characterised by a single ionization parameter, equation (3.5.16) can be implemented using the measured column densities in hydrogen and some arbitrary metal ion, together with the inferred ionization fractions in each, and assuming solar abundance ratios. When there is a range in ionization parameter, the situation is more complicated, since the measured column density in each ion will be the sum of the column densities arising in each region (each characterised by a different value of U). In this case, the measured column density in some ion M_i will be expressed as

$$N(M_i) = \sum_k N(M)_k f(M_i)_k, \tag{3.5.17}$$

where $N(M)_k$ and $f(M_i)_k$ are the k total column densities of element M and ionization fractions of ion M_i respectively, for a set of k ionization parameters. In the limit where there is a continuous distribution over ionization parameter, this becomes

$$N(M_i)_{\mathrm{obs}} = \int_{\xi_{\mathrm{min}}}^{\xi_{\mathrm{max}}} \frac{\mathrm{d}N(M)}{\mathrm{d}\xi} f(M_i)\, \mathrm{d}\xi, \tag{3.5.18}$$

where $\xi = \log U$.

To estimate total column densities in velocity components v_2 and v_4 (where we have the largest number of well-measured ions), we assume that all of the gas is photoionized, and plot the locus of points that mark the observed column densities in $\log U - \log N_{\rm H}$ space as shown in Figures 3.13 and 3.14, where the width of each contour indicates the 1σ uncertainty. The left-hand panels represent the case where the gas is photoionized by the 'UV peak' SED, with assumed metallicities $[M/{\rm H}] = -2, -1, 0$, and 1. The right-hand panels represent the equivalent scenario, but with the 'Soft X-ray peak' SED. Various zones within the absorbing region are physically characterised by the pair $(U, N_{\rm H})$ that best predict the observed column densities. Constraints on the ionization parameter are identified on the basis of column density ratios that are labelled similarly to those in Figure 3.10. Ratios that involve one ion having a lower limit on the column density give corresponding lower limits on the ionization parameter in Figures 3.13 and 3.14, plotted with vertical dashed lines and corresponding arrows. In velocity component v_2, we identify a minimum of two ionization components $(U, N_{\rm H})$ that can account for all of the observed column densities. Ionization parameters in these components are inferred from the ratios N(O IV)/N(Ar VIII) and N(Ne VIII)/N(Mg X). This conclusion comes with the caveat that we require there to be an under-abundance of nitrogen (and an over-abundance of sulphur, in the case of the UV peak SED) by factors of a few with respect to the solar values to explain why the O IV, N IV, S V and Ar VIII contours do not all cross in the $\log U$ range defined by N(O IV)/N(Ar VIII). In velocity component v_4, we require there to be at least three ionization components, with a similar caveat on the assumed overabundance of sulphur. This result contrasts with that from some previous, similar studies, that find a maximum of two discrete ionization components are required to adequately fit the data (e.g. Moe et al., 2009; Edmonds et al., 2011; Borguet et al., 2012; Arav et al., 2013). In this case, the ionization components inferred are no longer robust, since there are non-negligible fractional abundances of e.g. O IV and N IV in both of the lower ionization components. The limiting scenario, is one where there exists a continuous distribution of ionization parameters

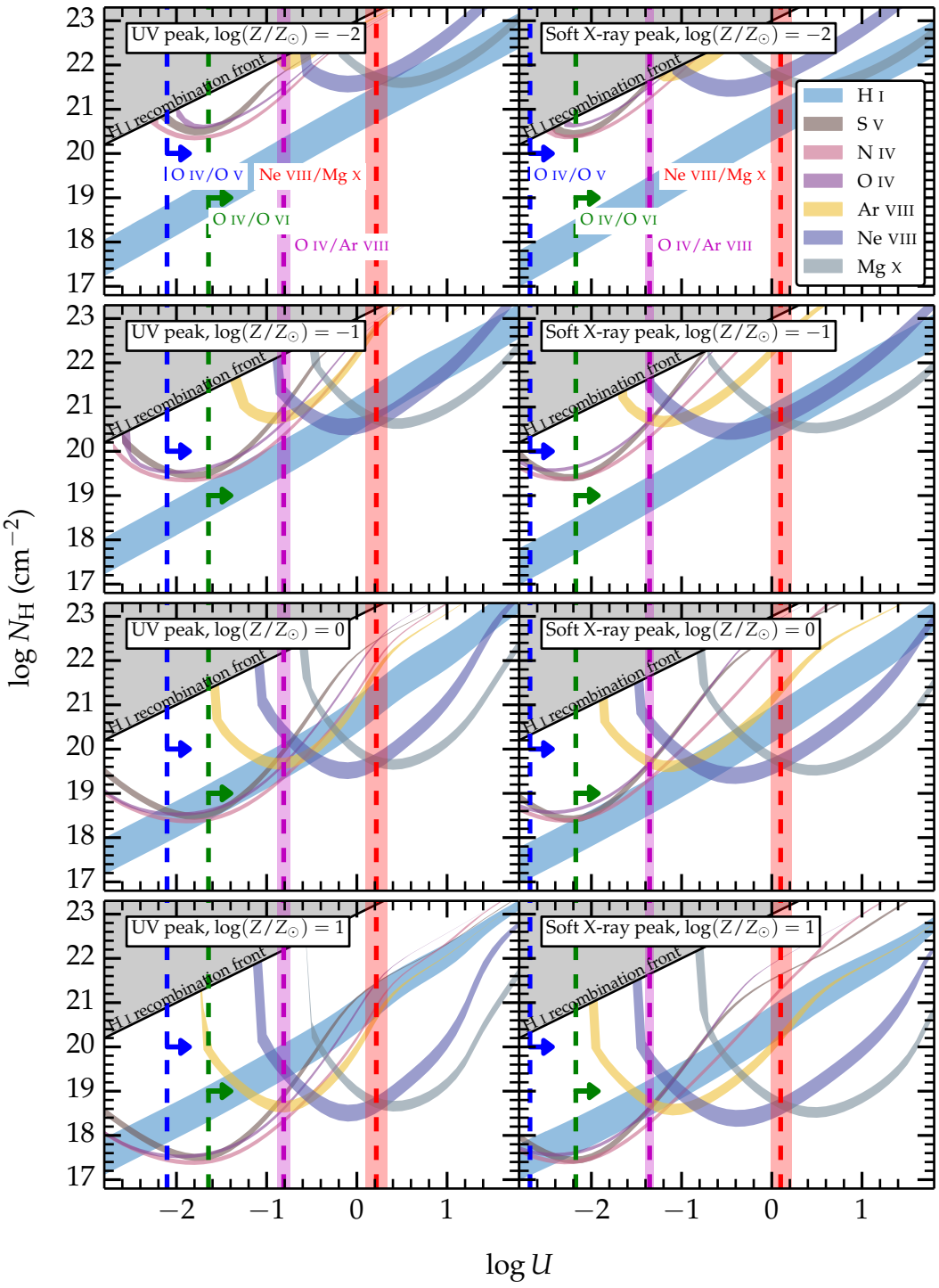

Figure 3.13: Ionization solutions for component v_2 in the case of the 'UV peak' incident SED (left panels) and the 'Soft X-ray peak' SED (right panels) for a range of metallicities. Coloured regions represent the values of $\log U$ and $\log N_{\rm H}$ that predict the $\pm 1\sigma$ bounds on the column densities for each labelled ion. Vertical bars indicate the plausible range in U for each of the ionization components identified by the column density ratios labelled. The crossing points between contours, defining the pair (U, $N_{\rm H}$), physically characterise these components. Dashed vertical lines and arrows represent lower limits on U.

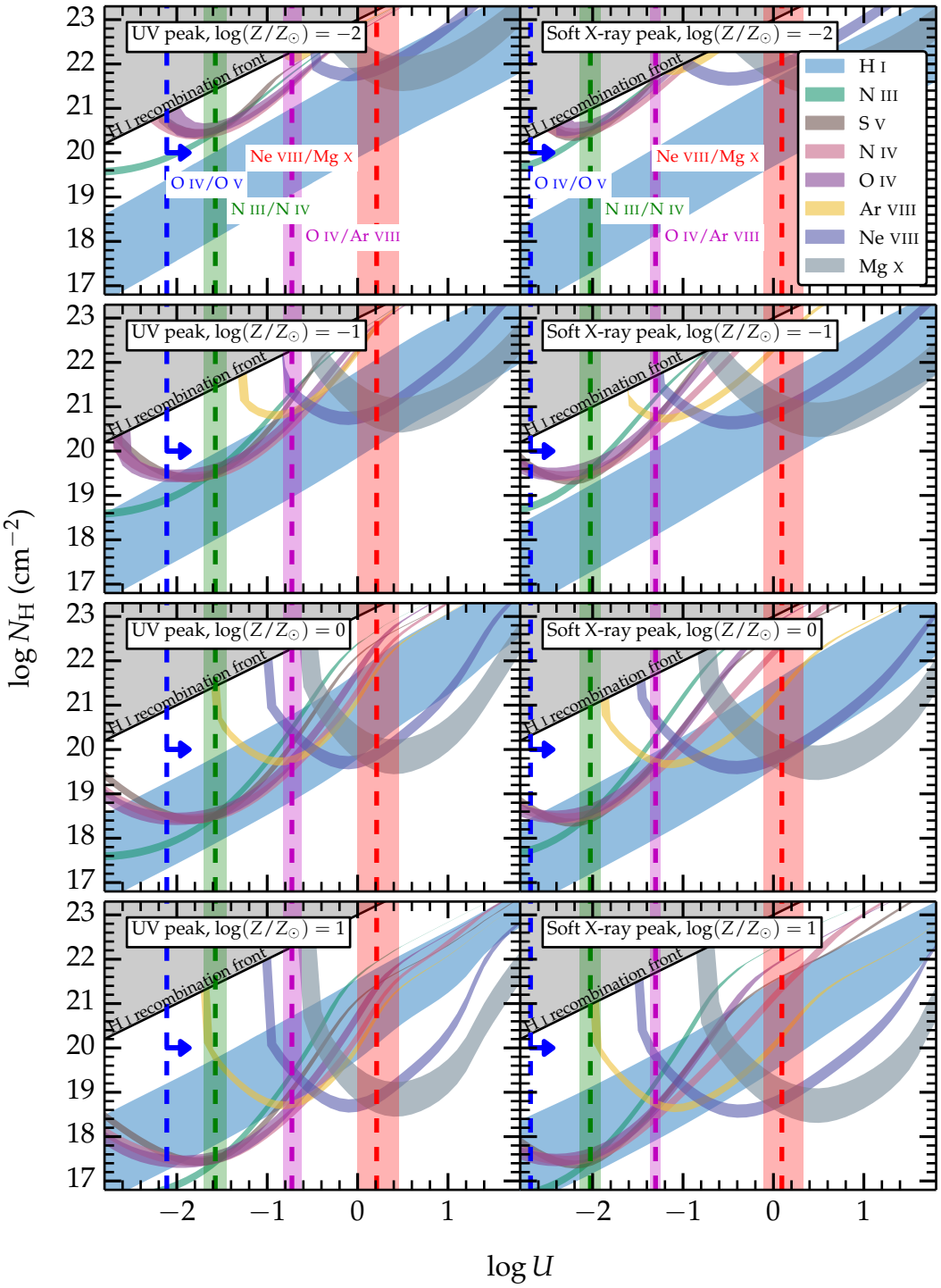

Figure 3.14: Ionization solutions for component v_4 in the case of the 'UV peak' incident SED (left panels) and the 'Soft X-ray peak' SED (right panels) for a range of metallicities. This figure has the same format as Figure 3.13.

through the absorbing region, which at a fixed distance R corresponds to a smooth variation in gas density. This scenario is described by equation (3.5.18). To simplify the problem, we assume that a (close-to) continuous distribution in U exists in the AAL region, and find an approximate solution to equation (3.5.18) assuming that each ion forms largely at the peak in its fractional abundance, $f(M_i)_{\rm max}$. In this case, equation (3.5.18) becomes

$$N(M) \simeq \frac{N(M_i)}{f(M_i)_{\rm max}}, \tag{3.5.19}$$

and we can read off a corresponding ionization parameter from Figure 3.10.

Assuming that most of the gas hosting H I also hosts most of the low ionization species in our sample (e.g. N III, N IV, O IV and S V), we can identify peak fractional abundances of these ions with fractional abundances of H I in Figure 3.10, and find the metallicity of the gas using equation (3.5.16). Incorporating uncertainty in the SED, we conservatively conclude that $0 \lesssim$ [O/H] $\lesssim 1$. We find this metallicity to be representative across all velocity components, and find the same result for [N/H] and [S/H]. Peak fractional abundances correspond to column density loci minima in Figures 3.13 and 3.14. Splitting the $\log U$ range in half at $\log U = -1$, it is clear that the high ionization gas traced by e.g. Ne VIII and Mg X has a factor of ~ 10 higher contribution to the total column density compared to the low ionization gas traced by e.g. N III and O IV. Constructing plots like these for all other velocity components (not shown), we conservatively conclude that, for each of the velocity components $v_1 - v_7$, the total hydrogen column density is $10^{17} \lesssim N_{\rm H} \lesssim 10^{18.5}$ cm^{-2} in the low ionization gas, and $10^{18.5} \lesssim N_{\rm H} \lesssim 10^{20}$ cm^{-2} in the high ionization gas. In velocity component v_8, we detect only O IV, O V, and N IV absorption, and the total hydrogen column density through this region is, conservatively, $10^{16.5} \lesssim N_{\rm H} \lesssim 10^{17.5}$ cm^{-2}.

3.5.5 Distance and size constraints

Given the estimate on $n_{\rm H}$, combined with information on U and $N_{\rm H}$, we can put a constraint on both the distance to the absorbing clouds from the QSO, and their

geometry. To determine the distance, R, we simply rearrange Equation (3.5.15) to obtain

$$R = \sqrt{\frac{Q_{\rm H}}{4\pi c n_{\rm H} U}}, \tag{3.5.20}$$

where

$$Q_{\rm H} \equiv \int_{\nu_{\rm LL}}^{\infty} \frac{L_\nu}{h\nu} d\nu \tag{3.5.21}$$

expresses the rate of emission of photons having energies sufficient to ionize hydrogen. Incorporating the uncertainty in $Q_{\rm H}$ (due to uncertainties in the SED), and conservatively estimating from Figure 3.10 that absorption due to O IV and O IV* is expected to arise in gas with $-2.4 \lesssim \log U \lesssim -2.0$ (from peak fractional abundances), we find that $2.3 \lesssim R \lesssim 6.0$ kpc. Since the velocity structure is consistent with being the same across all ions, this distance is likely to be representative for the AAL region as a whole. Therefore, if all of the gas were photoionized, given the expression for U in equation (3.5.15), we should expect clouds with differing densities. The ionization parameter is around two orders of magnitude different between the gas hosting the majority of O IV and the gas hosting the most highly ionized species (Ne VIII, Mg X), so this gas should have a characteristic density around two orders of magnitude smaller, i.e. $n_{\rm H} \sim 10$ cm^{-3}.

We can now estimate the absorption path length through the AAL region as $l_{\rm abs} = N_{\rm H}/n_{\rm H}$. Adopting $n_{\rm H} \sim 10^3$ cm^{-3} through the least ionized gas ($10^{17} \lesssim N_{\rm H} \lesssim 10^{18.5}$ cm^{-2}) and $n_{\rm H} \sim 10$ cm^{-3} through the most highly ionized gas we detect ($10^{18.5} \lesssim N_{\rm H} \lesssim 10^{20}$ cm^{-2}), we derive characteristic absorption path lengths of $10^{-4.5} \lesssim l_\parallel \lesssim 10^{-3}$ pc and $0.1 \lesssim l_\parallel \lesssim 1$ pc respectively in each velocity component. Note that the absorption path length is even smaller for velocity component v_8. If we assume that the absorbing gas completely fills the volume it encompasses, then these path lengths are also representative of the cloud sizes along the line-of-sight. If instead there are many clouds contributing to the total column density in each component, then the cloud sizes will be *smaller*, and the volume they encompass *larger* than the absorption path length. Additionally, based on indications

in Section 3.4.1 that the QSO continuum is only partially covered by the absorbing clouds, the transverse sizes of these gas clouds are likely to be $l_{\perp} \lesssim 10^{-2.5}$ pc (Jiménez-Vicente et al., 2012).

3.6 Discussion and conclusions

The main results from the data analysis and photoionization/collisional ionization equilibrium models are as follows:

1. The gas traced by the AALs is predominantly photoionized.

2. Under photoionization equilibrium, multiple ionization parameters are required to reproduce the column density ratios seen in the data.

3. Based on the observed column densities and ionization fractions implied from characteristic ionization parameters, incorporating uncertainties in the shape of the QSO SED, the gas metallicity is conservatively $0 \lesssim [\mathrm{O/H}] \lesssim 1$.

4. Given the range in possible gas metallicity, the total hydrogen column density in each velocity component is $10^{17} \lesssim N_{\mathrm{H}} \lesssim 10^{18.5}$ cm^{-2} in the least ionized gas (with slightly smaller values for velocity component v_8) and $10^{18.5} \lesssim N_{\mathrm{H}} \lesssim 10^{20}$ cm^{-2} through the most highly ionized gas we detect.

5. Taking the column density ratio between O IV* and O IV, assuming the fine structure excited states are populated mostly due to collisions with electrons, the total hydrogen density in the gas traced by these ions is found to be $\log(n_{\mathrm{H}}/\mathrm{cm}^{-3}) \sim 3$ for solar metallicity.

6. Given the total hydrogen density, and the plausible range in ionization parameter for the gas traced by O IV and O IV*, the distance to the absorbing clouds from the centre of the QSO is found to be $2.3 \lesssim R \lesssim 6.0$ kpc. An empirically identified, shared velocity structure amongst all ions, suggests this distance determination is likely to be representative for the AAL region as a whole.

7. Under photoionization equilibrium, the total hydrogen density in the most highly ionized AAL gas is found to be two orders of magnitude lower than that implied for the least ionized gas.

8. The ratio $N_{\rm H}/n_{\rm H}$ sets limits on the absorption pathlength through the least and mostly highly ionized regions we detect of $10^{-4.5} \lesssim l_{\parallel} \lesssim 10^{-3}$ pc and $0.1 \lesssim l_{\parallel} \lesssim 1$ pc respectively in each velocity component.

9. Covering fractions less than unity (in all cases where they can be reliably measured), suggest that the continuum region is only partially covered, requiring clouds with transverse sizes $l_{\perp} \lesssim 10^{-2.5}$ pc.

In summary, the analysis of the previous sections has revealed the presence of metal-enriched (to at least solar), highly ionized gas clouds a few kpc from the centre of Q0209 that are likely to be very small (sub-pc scale). In the following sections we place these results in a wider context, and speculate on the origins and fate of the absorbing gas. For simplicity, we shall speak of two co-spatial, photoionized regions: a low ionization region with $\log U \lesssim -1$, and a high ionization region with $\log U \gtrsim -1$.

3.6.1 Gas structure and dynamics

A redshift measurement for Q0209 of $z_{\rm QSO} = 1.13194 \pm 0.00001$ implies that the AAL gas is mostly outflowing from the QSO with velocities up to ~ 400 km s^{-1} (see Figure 3.8). This is unusually small, compared to the majority of the AALs and BALs in the literature with high ionization species such as Ne VIII and Mg X, which are typically outflowing with velocities closer to a few thousand or few tens of thousand km s^{-1} (e.g. Hamann et al., 1995; Telfer et al., 1998; Petitjean & Srianand, 1999; Arav et al., 1999; Muzahid et al., 2012, 2013), although see Hamann et al. (2000) for a more similar example. If we assume that the gas is moving with a constant radial velocity v, and originates close to the SMBH, then the time-scale for reaching its current radius, R, is at least

$$t \approx 10^7 \left(\frac{R}{2.3 \text{ kpc}}\right)\left(\frac{200 \text{ km s}^{-1}}{v}\right) \text{ yrs.} \tag{3.6.22}$$

Different velocity components in the AAL gas are moving at different speeds, so the overall region should possess an appreciable radial thickness after a time t, even though we derive densities and ionization parameters that are consistent with one another across the different velocity components.

Given the small cloud sizes in the low ionization gas, a key question is how long they are expected to survive. The free-fall time-scale for these clouds can approximately be expressed as

$$t_{\rm ff} \equiv \frac{1}{\sqrt{G\rho}} \sim 2.4 \times 10^{15}\ {\rm s} \left(\frac{n_{\rm H}}{\rm cm^{-3}}\right)^{-1/2}, \tag{3.6.23}$$

where G is the gravitational constant, ρ is the gas density, assuming that all of the mass is baryonic, and setting the mass fraction in helium to 0.28 (assuming solar abundances). In addition, for a characteristic cloud size l, the sound crossing time in a highly ionized plasma can be approximated by

$$t_{\rm sc} \equiv \frac{l}{c_{\rm s}} \sim 2.1 \times 10^{15}\ {\rm s} \left(\frac{l}{\rm kpc}\right) T_4^{-1/2}, \tag{3.6.24}$$

where $c_{\rm s}$ is the sound speed in an ideal monatomic gas, and $T = T_4 \times 10^4$ K (e.g. Schaye, 2001). For a stable cloud in hydrostatic equilibrium, $t_{\rm sc} \sim t_{\rm ff}$. We take a value of $n_{\rm H} = 10^3$ cm^{-2}, and a value of $l = 10^{-6}$ kpc (assuming $l_{\parallel} \approx l$). The photoionization models indicate that $T_4 \approx 2$ in this gas, and so we find $t_{\rm ff} \sim 7.6 \times 10^{13}$ s $\gg t_{\rm sc} \sim 1.5 \times 10^9$ s. This implies that the clouds will expand on the sound crossing time-scale, so they should have lifetimes of $\lesssim 100$ years. This is considerably less than the characteristic flow time in equation (3.6.22), and so the probability of observing these clouds at their implied distance from the QSO is extremely small in this case.

The analysis presented above poses a problem, which may be overcome if the clouds are being held in pressure equilibrium. This may be a thermal pressure equilibrium with higher temperature, lower density, more highly ionized gas, equivalent to the statement $n_{\rm H1}T_1 = n_{\rm H2}T_2$, where $n_{\rm H1}$, T_1 and $n_{\rm H2}$, T_2 are the total hydrogen densities and temperatures of the low and high ionization regions

respectively. From Section 3.5, under photoionization equilibrium, we found that $n_{\rm H2} \sim 10$ cm^{-2}, and these models also indicate that $T_2 \approx 6 \times 10^4$ K. In this case, $n_{\rm H1}T_1 > n_{\rm H2}T_2$, and the high ionization gas cannot pressure support the low ionization gas. If the former is collisionally ionized, we now have temperatures that differ by more than an order of magnitude. Densities in the high ionization region may be low enough to allow for pressure support. Nevertheless, the high ionization gas itself, accounting for the possibility that it is photoionized, should have a lifetime $\lesssim 10^5$ years, which is still short enough to suggest that this gas may also require pressure support from even more highly ionized gas that we do not detect in the UV, and which would require larger total column density, higher temperature, and lower density.

Massive galaxies are expected to host hot gas coronae, well within the implied location of the AAL region, with $T \sim 10^6$ K and $n_{\rm H} \sim 10^{-2}$ cm^{-2} (e.g. White & Frenk, 1991; Fukugita & Peebles, 2006). Pressure from this external medium, together with additional pressure support from magnetic confinement (de Kool & Begelman, 1995) may help to alleviate the problems outlined above, although pressure supporting gas with varying internal pressure is clearly a complex issue. We note that the analysis above does not incorporate the effects of turbulence, which is almost certainly present given the observed line widths (see Table 3.4). In addition, gas outflowing from a QSO will likely encounter the interstellar medium (ISM) of the host galaxy on its journey out into the halo. At supersonic velocities, shocks will likely occur at the interface between the outflowing gas and the ISM, heating the gas close to this interface. The resulting mix of hot and cool gas creates instabilities that can destroy the clouds before they reach the halo (see, for example, arguments in Faucher-Giguère et al., 2012). These authors suggest an alternative scenario, in which small clouds may be formed in-situ from moderately dense ISM clouds within hot, recently shocked gas. These clouds become shredded by a passing blast wave, and gain momentum from an accompanying shock. The resulting 'cloudlets' in this model have sizes and densities comparable to those derived here, and can possess a range of velocities that may explain the multi-component velocity structure in the absorption profiles. This in-

situ condensation scenario may also be consistent with the observation of dense molecular and atomic gas observed in other QSO outflows (e.g. Feruglio et al., 2010; Sturm et al., 2011; Aalto et al., 2012; Maiolino et al., 2012; Cicone et al., 2014), which are also predicted by some outflow models and simulations (e.g. Zubovas & King, 2014; Costa et al., 2015).

Models such as these may offer a more promising route to explain the structure and dynamics of AALs with properties (density, cloud size, velocity structure, distance from the QSO) similar to those found in Q0209 (e.g. Petitjean & Srianand, 1999; Hamann et al., 2000; Edmonds et al., 2011; Borguet et al., 2012; Arav et al., 2013; Muzahid et al., 2013). Any viable model must additionally reproduce the covering fractions seen in the present data. Covering fractions less than unity, and with little variation, are seen in ions spanning a range in ionization potential from a few tens to a few hundreds of eV, tracing gas with more than one possible ionization mechanism. Vastly differing absorption path lengths through the AAL region, as hinted at in the analysis of the previous sections, make it very difficult to account for the near constancy in covering fraction across all ions using simple geometrical models. We also note that these results differ from e.g. Hamann et al. (2000) and Borguet et al. (2012), who find more complete coverage in high-ionization UV transitions compared to those at lower ionization potentials. Covering fractions less than unity across our sample also go against general trends for more complete coverage with lower outflow velocities, as identified in the COS sample presented by Muzahid et al. (2013). It is therefore clear that simple trends such as these may not produce robust predictions for individual systems, which further highlights the apparent complexity in these absorbers.

3.6.2 Are the AAL clouds out of equilibrium?

Up to this point, our analysis and discussion has assumed that the AAL clouds are in ionization equilibrium. However, in general, absorbers may be out of equilibrium when close to an AGN due to recombination time-scales that can be long compared to typical AGN lifetimes and duty cycles (e.g. Krolik & Kriss, 1995; Nicastro et al., 1999; Arav et al., 2012; Oppenheimer & Schaye, 2013a,b). The re-

sulting recombination lag can lead to situations where high ionization stages like O VI, Ne VIII, and Mg X are enhanced relative to the expectation from equilibrium models. We examine these issues here.

We define the the photoionization rate, Γ_{M_i} (s^{-1}), for a given ion M_i as

$$\Gamma_{M_i} \equiv \int_{\nu_{0,M_i}}^{\infty} \frac{4\pi J_\nu}{h\nu} \sigma_{M_i}(\nu)\, \mathrm{d}\nu. \tag{3.6.25}$$

Here ν is the frequency, ν_{0,M_i} is the ionization frequency, J_ν is the intensity of the QSO radiation field (erg s^{-1} cm^{-2} Hz^{-1} sr^{-1}), σ_{M_i} is the photoionization cross-section, and h is Planck's constant. The recombination rate (s^{-1}) into an ion M_i is given by

$$R_{M_i} \equiv \alpha_{M_i} n_e, \tag{3.6.26}$$

where n_e is the electron number density, and α_{M_i} is the temperature dependent recombination rate coefficient (cm^3 s^{-1}) for that ion. Finally, the collisional ionization rate for an ion M_i is

$$C_{M_i} \equiv \beta_{M_i} n_e, \tag{3.6.27}$$

where β_{M_i} is the collisional ionization rate coefficient (cm^3 s^{-1}), which is also temperature dependent. Neglecting Auger ionization and charge transfer, the population in an ion M_i is then

$$\frac{\mathrm{d}n_{M_i}}{\mathrm{d}t} = -n_{M_i}(\Gamma_{M_i} + R_{M_{i-1}} + C_{M_i}) + n_{M_{i+1}} R_{M_i} + n_{M_{i-1}}(\Gamma_{M_{i-1}} + C_{M_{i-1}}). \tag{3.6.28}$$

Now suppose that an absorber is in photoionization equilibrium, i.e. $\mathrm{d}n_{M_i}/\mathrm{d}t = 0$, at time $t = 0$, but there is a sudden change in the ionizing flux, such that $\Gamma_{M_i}(t > 0) = (1+\delta)\Gamma_{M_i}(t = 0)$, where $-1 \leq \delta \leq \infty$. Taking the collisional ionization rate to be negligible (a reasonable approximation for a photoionized plasma), it can then be shown that the e-folding time-scale for change in the ionic fraction is

given by

$$t_{\rm change} = \left[-\delta \alpha_{M_i} n_e \left(\frac{n_{M_{i+1}}}{n_{M_i}} - \frac{\alpha_{M_{i-1}}}{\alpha_{M_i}} \right) \right]^{-1}, \tag{3.6.29}$$

(Arav et al., 2012), where negative time-scales indicate a decrease in the ionic fraction, and positive time-scales indicate an increase. For changes in the ionizing flux within an order of magnitude ($0.1 < 1 + \delta < 10$), these time-scales are typically $\sim$ 10 years for the densities implied by the O IV* analysis of Section 3.5.1, assuming $T \sim 10^4$ K. Since the gas densities in the AAL region are much higher than those typical of the diffuse IGM and circumgalactic medium (CGM), these time-scales are much shorter than the typical AGN lifetime ($\sim$ Myr time-scales; e.g. Novak et al., 2011). Therefore, photoionization equilibrium might be a valid assumption in our case, so long as these time-scales are also short compared to the time-scale over which the QSO luminosity changes.

However, we must also consider the dynamical evolution of the AAL clouds. In the previous section, we found that unless the clouds traced by O IV are pressure supported by an external medium, they will expand on time-scales $\lesssim$ 100 years. If this process is occurring, then the clouds may be entering a non-equilibrium state due to a recombination lag. To determine whether or not this scenario is likely, we numerically solve the coupled, time-dependent ionization equations (equation (3.6.28)) for a set of elements using a 4th order Runge-Kutta method. We assume that the AAL region possesses a gas density of $n_{\rm H} = 10^3$ cm^{-3}, and is illuminated by the 'UV peak' SED at a distance of 2.3 kpc. Assuming a gas temperature $T = 10^4$ K, we then calculate recombination rate coefficients using the Badnell (2006) fits (assuming case A recombination), and collisional ionization rate coefficients using the data in Voronov (1997). We perform the integral in equation (3.6.25) using the same photoionization cross-sections as used in Cloudy c13.00. Equilibrium values of n_{M_i} are then calculated, assuming solar abundances. We assume these values to hold at a time $t = 0$. We next perturb the gas density such that $n(t > 0) = (1+\delta)n(t = 0)$, where $-1 \leq \delta \leq \infty$, identical to the case of flux changes considered above. We do this for four different values of $(1+\delta) = 0.1, 0.01,$

0.001, and 10. The latter is included for the interest of comparing both increasing and decreasing density. For example, if the AAL clouds are the result of shocked ISM clouds, we might expect them to be crushed prior to their subsequent expansion (e.g. Faucher-Giguère et al., 2012). The resulting time-dependent evolution in the number density of H I, O IV, O V, O VI, Ne VIII, and Mg X is shown in Figure A.2 in Appendix A.1.

Simple inspection of Figure A.2 indicates that, for a decrease of more than an order of magnitude in gas density, with the exception of H I, time-scales for reaching a new equilibrium are $>$ 100 years. Comparing this to the expansion time-scale, we conclude that non-equilibrium effects are important if the density in the AAL region is dropping by orders of magnitude. Time-scales are orders of magnitude shorter for an increase in density, and so we expect that a decrease in density should form the dominant contribution to any non-equilibrium behaviour. We note that the situation described is not physical – we expect a smooth change in density with time, not a step-function change. Nevertheless, lacking a physical motivation for the exact functional form, we present these results as an approximation. The non-equilibrium behaviour shown in Figure A.2 indicates considerable variation in the rate of change of ionic number density, with the more highly ionized species changing more rapidly. Under photoionization equilibrium, the AAL region was found to possess a range in density and absorption path-length covering orders of magnitude. This situation was required due to the co-spatial existence of ionic species tracing gas with multiple ionization parameters. If ionic number densities are changing at highly variable rates, then is possible to envisage scenarios whereby the fractional abundances of ions spanning a large range in ionization potential can all be high. We therefore speculate that it may be possible to find non-equilibrium models that reproduce all of the observed column densities in a single phase, with a single density and absorption path length. Such a scenario may be desirable, as it is in better concordance with the near constant covering fraction seen across these ions.

It is important to note that the calculations leading to inferred cloud lifetimes of $\lesssim$ 100 years assume that the cloud sizes, l, are approximately equal to the ab-

sorption path-length $l_{\|}$. However, in general, $l < l_{\|}$, which would make the cloud lifetimes even shorter. Scenarios where $l < l_{\|}$ have been put forward multiple times in the literature, typically to explain situations where the derived covering fractions in the data depend on velocity across the absorption profiles, and the ionization and/or true optical depth in the lines. This situation is referred to as inhomogenous partial coverage (e.g. de Kool et al., 2002; Hamann & Sabra, 2004; Arav et al., 2008), where there are many small clouds having a power-law dependence in optical depth across their transverse extent. Covering fractions that vary in the way just described are not found in our data. Nevertheless, we cannot rule out the possibility that there are multiple small clouds along the line-of-sight. In such a scenario, the case for non-equilibrium evolution in the ionic number densities becomes more compelling.

In summary, if the AAL region is not pressure supported, then explicit numerical calculations suggest that non-equilibrium effects may be important in these clouds. A cloud expansion time-scale of $\lesssim 100$ years is sufficiently short that these effects might be confirmed with repeat observations. These observations will be crucial in determining appropriate non-equilibrium models for the data.

3.6.3 The connection to associated X-ray absorption

We next consider the link between associated UV absorption and so-called 'warm absorbers', often characterised by both bound-bound and bound-free absorption in X-rays. A key question is whether or not this absorption is predicted by the UV absorption lines characterised here. In the most highly ionized gas, under photoionization equilibrium, the maximum predicted total hydrogen column density is $\log(N_{\rm H}/{\rm cm}^{-2}) \approx 20$, and the ionization parameter $\log U \approx 0.5$. We assume an upper limit on the gas metallicity of [O/H] ≈ 1. Explicit photoionization calculations using these parameters predict the presence of bound-bound transitions, but no significant bound-free absorption in X-rays. The same result is obtained in CIE calculations. In this respect, and also in terms of their relatively small velocity shift from the QSO, the AALs in Q0209 are similar to those reported in e.g. Hamann et al. (2000). If there is continuous X-ray absorption, it should be in much more

highly ionized gas with larger total column density. This gas may trace the bulk of an outflow that produces the gas condensations described in the previous section. However, in general, the gas giving rise to warm absorbers need not be co-spatial with UV and/or optical AALs, especially since these absorption systems arise in gas with a wide range of physical parameters (outflow velocities, ionization, covering fractions, distance from the QSO etc.; see, for example, Ganguly et al., 1999; Misawa et al., 2007; Nestor et al., 2008; Ganguly et al., 2013; Muzahid et al., 2013; Sharma et al., 2013). Nevertheless, it is intriguing that the AALs in Q0209 show nearly identical velocity component structure over $\sim$ 300 eV in ionization potential, suggesting that absorption lines from many ionization stages can indeed arise co-spatially. Simultaneous observations in the UV and X-rays will likely be required to gain deeper insights into the connection between warm absorbers and associated absorption lines (e.g. Di Gesu et al., 2013; Lee et al., 2013).

3.6.4 Outflow models

Before considering potential origins for the outflowing gas, we first perform a rough estimate of the mass and kinetic energy in the AALs. We assume the geometry of the outflowing gas traced by these data to be that of a thin, partially filled shell of material moving radially outwards from the centre of the QSO, the flux from which is modelled by the 'UV peak' SED. Under this geometry, the mass depends on the distance from the QSO ($R \sim 2.3$ kpc), the total column density through the AAL region (we find a value $N_{\rm H} \approx 2 \times 10^{20}$ cm^{-2}), and crucially, the global covering fraction, Ω, of the AAL gas, as opposed to the line-of-sight covering fraction that we measure. A rough estimate of this quantity comes from the incidence rate of associated absorption systems like that seen in Q0209. Muzahid et al. (2013) presented a sample of 20 quasar spectra observed with COS, from which associated absorbers were selected based on the presence of Ne VIII absorption. The incidence rate of these absorption systems was found to be $\sim$ 40%, and we consider this to be the closest representative sample in the literature at present, although the general conclusions below are not sensitive to reasonable estimations of this value. We therefore express the total gas mass in the UV AAL

region as

$$M \approx 6 \times 10^7 \left(\frac{\Omega}{0.4}\right) \left(\frac{N_{\rm H}}{2 \times 10^{20}\ {\rm cm}^{-2}}\right) \left(\frac{R}{2.3\ {\rm kpc}}\right)^2 M_\odot, \tag{3.6.30}$$

where we have assumed a mean molecular weight per H particle of $\mu_{\rm H} = 1.4$. The total kinetic energy in this gas is then

$$K \approx 2 \times 10^{55} \left(\frac{M}{6 \times 10^7 M_\odot}\right) \left(\frac{v}{200\ {\rm km\ s}^{-1}}\right)^2 {\rm erg}. \tag{3.6.31}$$

We can derive the average mass outflow rate $\dot{M}$, by dividing M by the dynamical time-scale R/v, and subsequently derive the kinetic luminosity of the gas as $\dot{K} = 0.5\dot{M}v^2$. This gives values of $\dot{M} \approx 5 M_\odot\ {\rm yr}^{-1}$ and $\dot{K} \approx 7 \times 10^{40}\ {\rm erg\ s}^{-1}$. It is instructive to bear in mind that, while these quantities are useful, there are good reasons to believe that the gas clouds traced by the AALs may not travel a distance R, and are instead accelerated close to their observed location (see Section 3.6.1).

We consider two primary sources for the gas flow generating the AALs: (i) line-driven winds, and (ii) supernova-driven winds. Line-driven winds, initially accelerated through radiation pressure from the SMBH accretion disc, are commonly invoked to explain the winds traced by BALs and AALs with velocities of a few 1000 km s^{-1}, and are a major source of energy injection into the ISM in models of AGN feedback (e.g. Scannapieco & Oh, 2004; Di Matteo et al., 2005; Hopkins & Elvis, 2010). Specifically, these models require kinetic luminosities to be $\dot{K} \gtrsim 0.1\%$ of the Eddington luminosity, $L_{\rm Edd}$. For Q0209, $\log(L/L_{\rm Edd}) \gtrsim 0$, which implies the kinetic luminosity in the AALs is at least two orders of magnitude below this level. In addition, models involving line-driven winds suggest they must be launched close to the SMBH at velocities of a few 100 km s^{-1} (e.g. Risaliti & Elvis, 2010), which is already the velocity of the AALs seen here at much larger distances. If the AALs are pressure confined in a line driven wind such as this, they must encounter significant drag from a surrounding medium to slow them down, or keep them from accelerating to much larger velocities. In the more likely case that the AALs are formed in-situ, a variety of velocities could in principle be observed.

For example, in the radiative shock model of Faucher-Giguère et al. (2012), AAL clouds will take a finite time to accelerate up to the velocity of the passing blast wave (see their equation (12)). Although the UV AAL clouds in Q0209 contribute only a small percentage of the kinetic luminosity required for significant AGN feedback into the surrounding ISM (and IGM), a much larger percentage may be carried by an associated, much more highly ionized warm absorber, with higher total column density, detectable as bound-free absorption in X-rays (e.g. Crenshaw et al., 2003; Gabel et al., 2005; Arav et al., 2007). We note that bound-free X-ray absorption is by no means ubiquitous, with some warm absorbers now being detected via absorption *lines* such as O VII. These can have $N_{\rm H}$ consistent with that seen in associated UV absorption lines (e.g. Di Gesu et al., 2013).

Supernova driven winds are thought to drive fountains of gas a few kpc into the haloes of massive galaxies, some of which is then expected to fall back on ballistic trajectories (e.g. Bregman, 1980; Fraternali & Binney, 2006, 2008; Marinacci et al., 2010). We find that the distance, velocity and density of the AAL gas in Q0209 is typical of galactic winds (e.g. Veilleux et al., 2005; Creasey et al., 2013). The infalling velocity component v_8 also indicates that some of the AAL gas may be on a return trajectory back towards the disc of the QSO host. The time-scale derived in equation (3.6.22) is consistent with expected QSO lifetimes (e.g. Novak et al., 2011), so if the clouds are a result of supernova driven winds, it is possible these winds were launched during a starburst phase. If we assume that the starburst proceeded at $\sim 10 M_{\odot}$ yr^{-1}, and that this star formation results in one supernova per $100 M_{\odot}$ with energy $E_{\rm SN} \sim 10^{51}$ erg, then $\dot{E} \sim 10^{42}$ erg s^{-1}. Only a fraction of this energy will be converted into the kinetic energy powering an outflow, and so the kinetic luminosity we derive for the AALs in Q0209 may be consistent with this simple estimation.

At present, the data are consistent with both an outflow driven by AGN and starburst activity in Q0209. Future X-ray observations of this QSO may help to distinguish between these two possibilities. In particular, associated bound-free X-ray absorption tracing gas with high total column density and kinetic luminosity would favour an origin closely tied with the AGN.

Chapter 4

Galaxy surveys along QSO sight-lines

4.1 Overview

In our study on the relationship between the intergalactic medium (IGM) and galaxies, we treat the galaxies as single entities (we do not consider their constituent parts - bulge, disk, stellar and gas components etc.). Instead, we simply measure their redshifts and their globally averaged magnitudes. We also quantify the globally averaged star formation activity in the galaxies via their spectral features. We only consider galaxies with spectroscopic redshifts, because we need to put constraints on the IGM/galaxy association down to < Mpc scales.

This chapter is split into five sections. In Section 4.2, we summarise the galaxy sample constructed for this work, and describe the rationale behind it. In Section 4.3, we briefly describe the process of obtaining spectroscopic redshifts, and present a redshift classification scheme that is unified across our sample. In Section 4.4, we describe the processes involved in constructing global astrometry and photometry solutions from combined data sets. In Section 4.5, we describe the measurement of spectral features. Finally, in Section 4.6, we explain how these measurements are transformed into an indication of star-formation activity.

4.2 The galaxy sample

The galaxy data is obtained from a number of different instruments and surveys. We include data collected by our own collaboration from the Deep Imaging Multi-Object Spectrograph (DEIMOS), Gemini Multi-Object Spectrograph (GMOS), Canada-France-Hawaii Telescope (CFHT) multi-object spectrograph and Very Large Tele-

scope (VLT) Visible Multi-Object Spectrograph (VIMOS) (hereafter, T14 and T14-Q0107; Morris & Jannuzi, 2006; Tejos et al., 2014). We make use of the Sloan Digital Sky Survey (SDSS) (Abazajian et al., 2009), 2dF Galaxy Redshift Survey (2dFGRS) (Colless et al., 2001), Galaxy and Mass Assembly (GAMA) survey (Driver et al., 2011), VLT VIMOS Deep Survey (VVDS) (Le Fevre et al., 2005) and VIMOS Public Extragalactic Redshift Survey (VIPERS) (Guzzo et al., 2014). We also include data from the Las Campanas Observatory (LCO)/Wide Field Reimaging CCD Camera (WFCCD) galaxy survey of 20 fields surrounding UV-bright quasars (QSOs) (hereafter, P11; Prochaska et al., 2011a), galaxy data around PKS 0405−123 presented in Johnson et al. (2013); hereafter, J13, and galaxy data around HE 0226−4110 and PG 1216+069 presented in Chen & Mulchaey (2009); hereafter, C09. The latter two surveys made use of the Inamori-Magellan Areal Camera & Spectrograph (IMACS) and Low Dispersion Survey Spectrograph 3 (LDSS3) at LCO.

Together, these surveys cover regions close to all of the QSO sight-lines used to characterise the IGM (see Chapter 2, Section 2.3). Some were conducted for the primary purpose of mapping galaxies close to a particular QSO sight-line, while others serendipitously cover regions where there are bright QSOs with Hubble Space Telescope (*HST*) spectroscopy. Those that fall in the latter category are the large SDSS, 2dFGRS, GAMA, VVDS, and VIPERS surveys. For SDSS, we adopt just those galaxies in the main sample, i.e. SDSS-I/II (see Abazajian et al., 2009, for details). We restrict our combined galaxy sample to 4×4 square degree fields (0.00487 steradians) centred on each QSO.[1] This means that we can sample galaxy-absorber pairs to transverse separations of $\sim$ 15 comoving Mpc at the median redshift of our sample, $\sim$ 10 comoving Mpc at $z \sim 0.07$, and $\sim$ 1 comoving Mpc at $z \sim 0.005$. We discard all objects with $z < 0.005$, regardless of their classification, on the basis that they may be stars. Some fields are made larger by virtue of there being more than one QSO that inhabits a particular 4×4 square degree region.

We summarise our combined galaxy sample in Table 4.1. As an indication of survey depth, we list the median redshift for each survey, and the 95th percentile

[1]Note that a number of surveys cover areas of sky that are smaller than this.

Table 4.1: The galaxy sample

Survey	$N_{\rm galaxies}$[1]	$z_{\rm median}$[2]	z_{95}[3]	$m_{\rm limit}$[4]	Reference
SDSS	41 342	0.10	0.19	$r < 17.77$	Abazajian et al. (2009)
2dFGRS	10 643	0.11	0.22	$b_{\rm J} < 19.45$	Colless et al. (2001)
GAMA	8636	0.22	0.40	$r < 19.8$	Driver et al. (2011)
VVDS	18 181	0.58	1.07	$I < 22.5$	Le Fevre et al. (2005)
VIPERS	24 183	0.70	1.06	$I < 22.5$[a]	Guzzo et al. (2014)
T14	1049	0.43	0.93	$R < 23.5$[b]	Tejos et al. (2014)
T14-Q0107	962	0.55	1.07	various[c]	Tejos et al. (2014)
P11	900	0.16	0.36	R < 20[d]	Prochaska et al. (2011a)
C09	810	0.36	0.64	R < 22	Chen & Mulchaey (2009)
J13	443	0.41	0.81	R < 23	Johnson et al. (2013)
ALL	**107 149**	**0.19**	**1.00**	–	–

[1] Number of galaxies with spectroscopically confirmed redshifts (not labelled 'c' - see section 4.3 for details).
[2] Median redshift for the survey.
[3] The 95th percentile of the redshift distribution.
[4] Magnitude limit for the survey.
[a] Colour cuts also applied.
[b] Priority given to objects with $R < 22.5$.
[c] VIMOS: $R < 23$, priority given to objects with $R < 22$. DEIMOS: $R < 24.5$, priority given to brighter objects, colour cuts also applied. GMOS: Top priority given to objects with $R < 22$, second priority given to objects with $22 < R < 23$, last priority given to objects with $23 < R < 24$. CFHT: $R < 23.5$ (indicative only).
[d] $R < 19.5$ for some fields.

of the redshift distribution, which we denote z_{95}. This is more informative than the maximum of the redshift distribution, as many surveys show long tails to high redshift due to the presence of luminous active galactic nuclei (AGN). We also list the magnitude limit for each survey, which in many cases is only indicative (see the table footnotes for more details). There are 107 149 galaxies in our combined sample, which has a median redshift of 0.19. In Table 4.2, we summarise the QSO sight-line fields. We list the number of QSOs in each field and give an indication of the area and comoving volume covered by each. For the latter, we define the edge of the volume by the minimum of $(z_{\rm QSO}, z_{95})$, where $z_{\rm QSO}$ denotes the maximum QSO redshift for the field.

The following sections describe the analyses performed on the 1D extracted galaxy spectra and photometric parent samples. Much of this analysis builds on that already presented in Tejos et al. (2014). For the SDSS, 2dFGRS, VVDS and VIPERS surveys, and all galaxy data presented in Chen & Mulchaey (2009), Prochaska et al. (2011a) and Johnson et al. (2013), we work from the catalogued

Table 4.2: QSO sight-line fields

Field name	N_{QSO}	Area (sr)[a]	V_c (Gpc^3)[b]	Instrument/survey
J0005+1609	1	0.00221	0.181	SDSS
J0029+1316	1	0.00003	0.002	WFCCD
J0058−3606	1	0.00218	0.153	2dFGRS
J0110−0218	3	0.00004	0.049	CFHT, VIMOS, DEIMOS, GMOS
J0120−2821	2	0.00487	0.623	2dFGRS
J0159+0023	1	0.00309	0.215	SDSS
J0209−0438	1	0.00314	4.700	VIPERS
J0228−1904	1	0.00005	0.024	IMACS, LDSS3
J0407−1211	1	0.00029	0.176	WFCCD, IMACS, LDSS3
J0426−5712	1	0.00135	0.041	2dFGRS
J0559−5026	1	0.00004	0.002	WFCCD
J0809+4619	1	0.00487	0.436	SDSS
J0835+2459	1	0.00487	0.369	SDSS
J0847+3445	1	0.00487	0.059	SDSS
J0919+5521	1	0.00487	0.201	SDSS
J0956−0453	1	0.00332	0.210	2dFGRS
J0956+4115	2	0.00487	0.452	SDSS
J1005+0134	1	0.00082	1.082	SDSS, VVDS, VIMOS
J1007+2929	2	0.01097	1.042	SDSS
J1013+3551	1	0.00487	0.088	SDSS
J1014−0418	1	0.00487	0.049	2dFGRS
J1022+0132	1	0.00004	0.027	SDSS, VIMOS
J1031+5052	1	0.00487	0.467	SDSS
J1051−0051	1	0.00289	0.213	SDSS
J1058+3412	2	0.01097	1.010	SDSS
J1118+5728	2	0.01950	1.697	SDSS
J1119+2119	1	0.00363	0.291	SDSS, WFCCD
J1121+4113	1	0.01097	1.060	SDSS
J1131+3114	1	0.00487	0.502	SDSS
J1139+6547	1	0.00487	0.378	SDSS
J1226+0319	3	0.01950	3.638	SDSS, IMACS, LDSS3, WFCCD
J1232+2009	1	0.00487	0.057	SDSS
J1301+5902	2	0.00487	0.488	SDSS
J1305−1033	1	0.00004	0.007	WFCCD
J1309+0819	1	0.00487	0.309	SDSS, WFCCD
J1357+0435	1	0.00091	1.225	SDSS, VVDS
J1427+2348	1	0.00487	0.372	SDSS
J1437−0147	1	0.00487	1.646	GAMA
J1438−0658	1	0.00182	0.078	2dFGRS
J1442+3526	1	0.00487	0.088	SDSS
J1528+2825	1	0.00487	0.382	SDSS
J1555+1111	1	0.00487	0.577	SDSS, WFCCD
J1620+1724	1	0.00487	0.169	SDSS
J2009−4849	1	0.00004	0.001	WFCCD
J2132+1008	1	0.00296	0.035	SDSS
J2155−0922	1	0.00003	0.003	WFCCD
J2158−3013	1	0.00487	0.183	2dFGRS, WFCCD
J2218+0052	1	0.00109	1.464	SDSS, VVDS, VIMOS
J2254−1734	1	0.00122	0.015	2dFGRS
J2351−0109	1	0.00256	0.201	SDSS

[a] Field area, approximating the survey region as a rectangle.
[b] Comoving volume covered by the survey up to the minimum of (z_{QSO}, z_{95}).

magnitudes, redshifts, and spectral line measurements (where available).[2] GAMA galaxy catalogues used in this study are from phase II of the survey, and are not publicly available at the time of writing. For SDSS, we make use of the spectral line measurements presented in Brinchmann et al. (2004) (see Section 4.6 for details).

4.3 Galaxy redshifts

The majority of the galaxy redshifts in our VIMOS, DEIMOS, GMOS, and CFHT samples were obtained by cross-correlating galaxy, star, and QSO templates from SDSS[3] with each observed spectrum (see Morris & Jannuzi, 2006; Tejos et al., 2014, for a full description). Each galaxy was then assigned a quality flag to indicate the reliability of the assigned redshift. The scheme is designed as follows:

- *Secure* ('a'): At least three well-identified spectral features (emission or absorption lines) or two well identified emission lines.
- *Possible* ('b'): Only one or two spectral features.
- *Uncertain* ('c'): No clear spectral features.

(Tejos et al., 2014). Flag 'c' is typically raised for spectra with low signal-to-noise ratio (SNR), or due to an intrinsic lack of observable features at the instrumental resolution. We do not use these redshifts in any of the forthcoming analysis. For all other galaxy redshifts, we map the corresponding quality flags onto our scheme to ensure a unified definition for 'secure', 'possible', or 'uncertain' as follows.

In SDSS, we simply adopt all galaxies with a warning flag of 0 (indicating no warnings) as being secure (label 'a'), and flag all other redshifts as 'c' (see Stoughton et al., 2002, for details on SDSS flags).

The 2dFGRS scheme is defined in terms of absorption redshifts and emission redshifts separately. In brief, for absorption redshifts, a quality parameter Q_a is

[2]Catalogues for 2dFGRS, SDSS, VVDS, and VIPERS galaxies are obtained from http://www.2dfgrs.net, http://skyserver.sdss.org/casjobs, http://cesam.oamp.fr/vvdsproject/vvds.htm and http://vipers.inaf.it/rel-pdr1.html respectively. Catalogues from the analysis in Chen & Mulchaey (2009), Prochaska et al. (2011a), and Johnson et al. (2013) were obtained from http://vizier.cfa.harvard.edu/viz-bin/VizieR.

[3]http://www.sdss.org/dr7/algorithms/spectemplates/

defined in terms of a variable R, being the ratio of peak to noise in the cross-correlation with the best fitting template, as follows:

$$Q_a = \begin{cases} 4 & R > 5.0, \\ 3 & R > 4.5, \\ 2 & R > 4.0, \\ 1 & R > 3.5, \\ 0 & \text{otherwise}, \end{cases}$$

with a further requirement that $Q_a = 3$ and $Q_a = 4$ redshifts are obtained to within 600 km s^{-1} across four and six of the eight spectral templates respectively. For emission redshifts, the parameter Q_e is defined as

$$Q_e = \begin{cases} 4 & \text{three or more detected lines}, \\ 2 & \text{two lines, or one strong line}, \\ 1 & \text{one weak line}, \\ 0 & \text{no lines}. \end{cases}$$

The combined redshift quality flag, Q_b, is then determined as $Q_b = \max(Q_a, Q_e)$, unless the difference between the absorption and emission redshifts is < 600 km s^{-1}, in which case, $Q_b = \max(Q_a, Q_e, 3)$, or if $Q_a \geq 2$ and $Q_e \geq 2$ and the difference between absorption and emission redshifts is > 600 km s^{-1}, in which case $Q_b = 1$. An overall redshift flag, Q, is determined via human verification of the automated redshift measurement, with the option of manually fitting Gaussian lines to spectral features as a means to obtain the redshift. The scheme is then as follows:

$$Q = \begin{cases} 5 & \text{reliable redshift, high quality spectrum}, \\ 4 & \text{reliable redshift}, \\ 3 & \text{probable redshift}, \\ 2 & \text{possible, but doubtful redshift}, \\ 1 & \text{no redshift could be estimated}. \end{cases}$$

(see Colless et al., 2001, for a detailed description). We perform the mapping Q ≥ 4 → 'a', Q = 3 → 'b', Q < 3 → 'c'.

The GAMA scheme is virtually identical to the 2dFGRS scheme, and we adopt the same quality flag mapping (see Driver et al., 2011, for more details).

The VVDS scheme is defined according to the following numbering scheme: (0) no redshift (no features); (1) tentative redshift (weak features, continuum shape); (2) secure redshift (several features); (3) very secure redshift (strong spectral features); (4) completely secure redshift (obvious spectral features); (9) redshift based on on single secure feature. Added to this are the prefixes 1 and 2, to mean broad line AGN and secondary target respectively (see Le Fevre et al., 2005, for more details). We perform the following mapping:

- {4, 14, 24, 214, 3, 13, 23, 213} →'a',
- {2, 12, 22, 212, 9, 19, 29, 219} →'b',
- {1, 11, 21, 211, 0} →'c'.

The VIPERS scheme is identical to that of VVDS, but with the addition of a decimal fraction to each flag depending on the photometric redshift from the accompanying 5-band CFHT Legacy Survey (CFHTLS) photometry. If the spectroscopic redshift falls within the 1σ confidence interval on the photometric redshift, a value of 0.5 is added. If it falls within the 2σ confidence interval, a value of 0.4 is added. If it falls outside the 2σ confidence interval, a value of 0.2 is added, and when there is no photometric redshift, a value of 0.1 is added (see Guzzo et al., 2014, for more details). We adopt the same mapping as for VVDS, regardless of the added decimal fraction.

Redshifts for galaxies presented in Chen & Mulchaey (2009) and Prochaska et al. (2011a) are only provided where they are deemed reliable. We therefore label objects having an assigned redshift with flag 'a', and all other objects flag 'c'.

In Johnson et al. (2013), the redshift flagging scheme is defined (A) secure (≥ 2 features); (B) 1 feature; (C) observed but no features; and (N) not observed. We perform the mapping A→'a', B→'b', {C, N} →'c'.

4.4 Global astrometry/photometry solutions

In two fields, J1005−0134 and J2218+0052, VIMOS observations obtained by our collaboration supplement VVDS galaxy data at small angular separations from each QSO (Tejos et al., 2014). We have improved the photometric and astrometric calibration for these data as follows.

Astrometry and R-band photometry was originally obtained from the VIMOS pre-imaging data. However, these fields overlap with the VIRMOS deep imaging survey (Le Fevre et al., 2004), for which the astrometric and photometric calibration is superior. This photometric data set extends coverage to the B, V, and I bands, and forms the basis for target selection in the VVDS. We therefore set about matching these data sets to ensure global astrometric and photometric consistency across these fields. We made use of the SExtractor (Bertin & Arnouts, 1996) and Scamp (Bertin, 2006) software packages to automatically map galaxy positions from the VIMOS detector plane to world coordinates using sources detected in SDSS as a reference. This brought the astrometric solution to within one arcsecond of the VIRMOS deep imaging survey, which is below the typical seeing level. We then cross-matched the photometric catalogues, and calculated the mean R-band magnitude offset needed to bring the two into statistical agreement. We did this for each VIMOS quadrant separately. The typical offset was ~ 0.4 magnitudes. Not all sources could be matched to those from the VIRMOS deep imaging survey due to regions of the imaging for that survey that are poorly calibrated. For these sources we keep offset VIMOS R-band magnitudes, whereas elsewhere we assign the appropriate matched $BVRI$ photometry.

4.5 Spectral line measurements

For the VVDS and VIPERS surveys, no spectral line measurements or indications of spectral type are made available. We therefore performed our own analysis where possible, as a means to estimate the star-formation activity for the galaxies in these surveys (see section 4.6 for details). We also performed this analysis on the VIMOS, GMOS and CFHT data collected by our collaboration (Morris & Jan-

Table 4.3: Spectral index definitions

Index	Blue continuum (Å)	Line (Å)	Red continuum (Å)
[O II]	3655–3705	3708.5–3748.5	3750–3800
Hδ	4030–4080	4082.0–4122.0	4125–4170
Hγ	4230–4270	4321.5–4361.5	4365–4400
Hβ	4785–4820	4842.5–4882.5	5030–5100
[O III]	4785–4820	4988.0–5028.0	5030–5100
Hα + [N II]	6460–6520	6544.5–6584.5	6610–6670
[S II]	6640–6700	6713.0–6753.0	6760–6810

nuzi, 2006; Tejos et al., 2014). Originally, the spectral types for these galaxies were determined by assigning the spectral type of the best fitting template as part of the redshift determination process. We improve upon this by measuring spectral line fluxes, as described below.

For each galaxy spectrum, where spectral coverage, resolution and SNR allowed, we estimated the integrated fluxes and local continuum level around the [O II], Hδ, Hγ, Hβ, [O III], Hα, [N II], and [S II] emission lines. For this, we used the spectral line indicies defined in Table 4.3, and a direct integration over spectral pixels. The continuum level is obtained by iteratively clipping points 1.5σ below the estimated continuum in a manner similar to that described in Chapter 2, Section 2.5, taking a mean of the 'un-clipped' pixels either side of the line, and linearly interpolating between the points defined by these means. This makes the continuum estimate reasonably robust to underlying absorption, but it occasionally fails at the edges of some spectra where there is a loss in sensitivity, leading to poor flux-calibration, and a rapid fall-off in the continuum. The other main cause for continuum misplacement is the presence of occasional contaminating zero-orders[4] lying on top of the galaxy spectra, or regions of bad sky subtraction.[5] Note that we do not accurately remove the stellar continuum (including the underlying stellar absorption) in our procedure. This inevitably affects the reliability of the inferred emission line fluxes (the Balmer emission lines in particular), how-

[4]Crowding on the detector can often lead to zeroth-order spectra landing on regions inhabited by first-order spectra.

[5]A phenomenon known as 'airglow', caused by recombination of atoms and luminescence in the upper atmosphere, produces a large number of contaminating narrow emission lines in the spectra of astrophysical objects that cannot always be reliably subtracted.

ever our approach suffices for the purposes of splitting the galaxy sample into star-forming and non star-forming populations (see Section 4.6 for details).

The line indicies in Table 4.3 are optimised for spectra taken with VIMOS at a spectral resolution $R = 200$, appropriate for the VVDS and VIPERS surveys, and for the VIMOS data presented in Tejos et al. (2014). For the GMOS and CFHT data, we narrowed the line indicies to reflect the higher spectral resolution obtained by these instruments (see Tejos et al., 2014, and references therein for details). For integrated line fluxes detected above a 3σ significance threshold, we also attempt to fit Gaussian profiles, which are usually adopted in preference to the pixel measurements. We revert to the pixel measurements when the fitting routine returns a Gaussian with zero amplitude, indicating that the fit has failed. All Gaussian fits are performed with a χ^2 minimisation employing the Levenberg-Marquardt algorithm. The rest-frame standard deviation of each Gaussian line is bounded between $0.5\sigma_{\rm LSF}$ Å and $\sqrt{10^2 + \sigma^2_{\rm LSF}}$ Å, where $\sigma_{\rm LSF}$ is the standard deviation of the (assumed Gaussian) line-spread function (LSF), which helps to identify broad emission lines that are likely of AGN origin, and contaminating sky lines. We fit the [O II] doublet, Hδ, Hγ, and [S II] doublet lines separately. The [O II] doublet is not resolved by our spectra, so we fit it as a single line. For the [S II] doublet, which is marginally resolved, we tie the Gaussian standard deviations in the fit, and fix the line ratio [S II] λ6716 / [S II] λ6731 to the expected (but not fixed)[6] ratio of 1/1.4 (Osterbrock, 1989). We fit the Hβ/[O III] line complex simultaneously, tying together the Gaussian standard deviations, and fixing the [O III] λ4956 / [O III] λ5007 ratio to the expected value of 1/2.98 (Storey & Zeippen, 2000). We also fit the Hα/[N II] complex simultaneously, tying together the Gaussian standard deviations, and fixing the [N II] λ6548 / [N II] λ6583 ratio to the expected value of 1/2.95 (Osterbrock, 1989). The [N II] lines are barely resolved from the Hα line in our spectra, so we perform an alternative, single Gaussian fit to just the Hα line in every case. If this fit gives a smaller χ^2 value than the three-component fit, we assign the resulting Gaussian parameters to the Hα line, and report no [N II]

[6] Assumes a gas density and temperature.

measurements in these instances. Despite not having sufficient spectral resolution to properly resolve the [N II] components, in instances where these lines are strong, the resulting line profile has definite asymmetry, which motivates us to decompose the line profile. Uncertainties on the fitted Gaussian parameters are estimated by generating 100 Monte Carlo realisations of the data. For each realisation, we add a number to every pixel flux, randomly generated from a Gaussian distribution of values with standard deviation equal to its 1σ uncertainty. Each of these 100 realisations are then fit using the same procedure as in the nominal case, and the standard deviation over the resulting best-fit parameter values are taken as the 1σ uncertainty on the measurement.

To identify bad measurements in our galaxy spectra, we have devised a flagging scheme as follows:

- Flag (0): No warnings.
- Flag (1): Measurement may be affected by the OH forest[7] between 8600 and 8700Å.
- Flag (2): Line was fit with the maximum/minimum allowed Gaussian standard deviation.
- Flag (3): Line coincides with a region above a user-specified sky spectrum threshold.
- Flag (4): Line may be affected by the O_2 telluric absorption at $\sim$ 7600 Å.
- Flag (5): Bad continuum reduced χ^2 (> 10).
- Flag (6): No spectral coverage.

The quality of the sky subtraction in our spectra makes the line measurements reasonably robust to flag (1). Flag (2) is implemented to identify potential broad-line AGN and contaminating sky lines. Flag (3) is mainly implemented to eliminate contaminating zero-orders. We found that the VIMOS ESOREX pipeline reduction

[7]A well known region of densely-packed airglow emission lines, mostly attributable to hydroxyl ions, that are not always effectively subtracted.

software often incorrectly identifies zero orders offset from the galaxy spectrum of interest and tries to correct for them, leaving deep, artificial absorption features in the extracted 1D galaxy spectra. Nevertheless, these appear as broad spikes in the extracted 1D sky spectra, and can be identified by adopting a threshold sky value. Flag (4) is implemented because most of our spectra are not corrected for the O_2 telluric, and even in spectra that are corrected for this contaminating feature, the correction is highly uncertain due to the narrow 'picket-fence' nature of the absorption. Flag (5) is implemented to pick up line measurements that are marred by bad continuum estimation. We allow for relatively high reduced χ^2 values in view of occasional absorption that raises the value of this statistic even for reasonable continuum estimations. Flag (6) is implemented to identify lines not measured due to insufficient spectral coverage. We enforce coverage across the entire region defined by each of the line indicies in Table 4.3. Line measurements that do not raise any of the aforementioned flags are assigned flag (0), to indicate that there are no warnings.

In all of our spectra, we reject measurements that raise flags 2, 4, 5, and 6. For VIMOS spectra reduced with ESOREX, we additionally reject measurements that raised flag 3. In practice, this flag is reserved for those spectra only.

4.6 Star formation activity

For the VVDS and VIPERS surveys, and our VIMOS, GMOS, and CFHT data, we use the spectral line measurements described in the previous section to split the sample of galaxies in terms of their star formation activity. We also do the same using a very similar set of line measurements provided by the GAMA survey team (see Hopkins et al., 2013, for a description). We aim simply to define galaxies as 'star-forming', 'non star-forming', 'AGN dominated', or 'unclassified'. Although we could have calculated star formation rates for many of our galaxies using standard procedures (e.g. Kennicutt, 1998a; Moustakas et al., 2006), estimating K-corrections at redshifts $\gtrsim 0.5$ becomes increasingly uncertain, and in general we lack a homogeneous set of multi-band photometric measurements across our

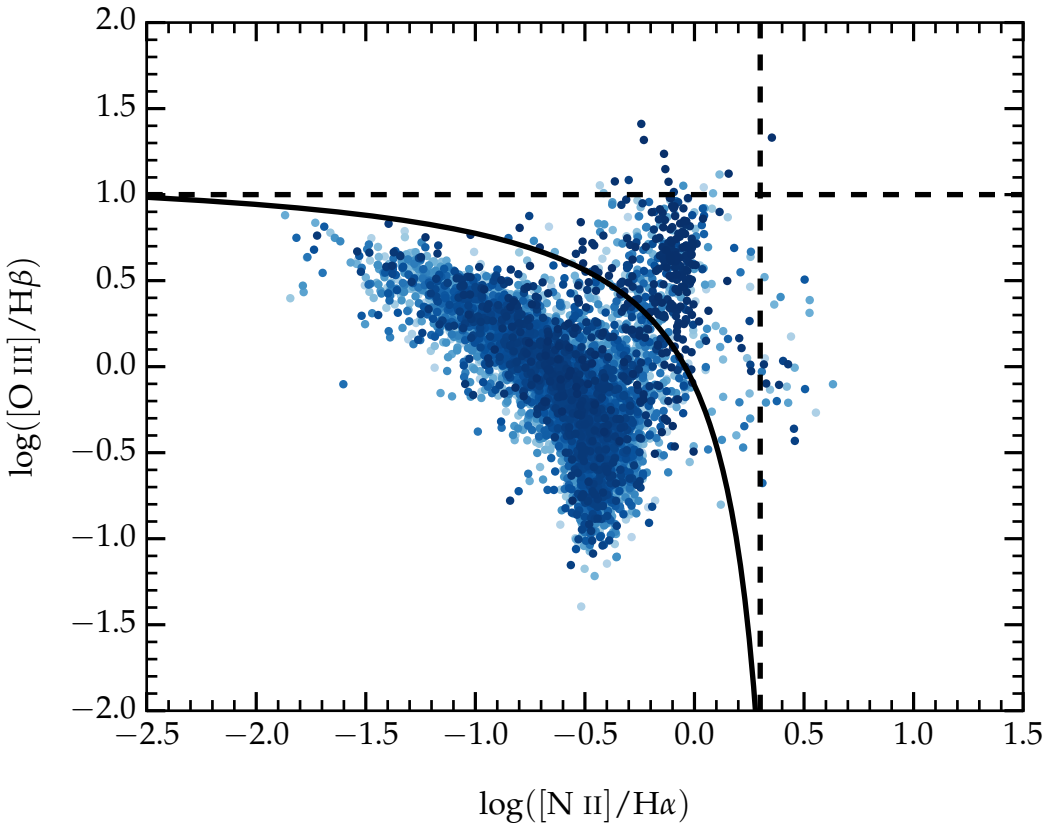

Figure 4.1: The distribution of galaxies in our sample on a BPT line-ratio diagram. The solid line indicates the discrimination line from Kewley et al. (2001), separating star-forming galaxies from AGN. The dashed lines indicate the single-line ratio diagnostics we also employ to identify AGN when two line ratios are not available.

sample to allow for a consistent approach. In any case, simply splitting our sample purely on the basis of spectral line measurements suffices for our requirements. We adopt a very similar prescription to that outlined in Brinchmann et al. (2004), whose classification scheme was applied on the SDSS galaxies in our sample.

First, we attempt to identify galaxies whose spectra are dominated by an AGN component. A number of broad-line AGN are already identified on the basis of their redshift determination via cross-correlation with AGN templates (Tejos et al., 2014). For the remaining galaxies, we perform a classification on the basis of a Baldwin, Phillips & Terlevich (1981, hereafter BPT) diagram, shown in Figure 4.1. Kewley et al. (2004) performed spectral energy distribution (SED) and photoionization modelling to find a theoretical discriminating line between star-forming galaxies and AGN on a BPT diagram of log([N II]/Hα) versus log([O III]/Hβ).

This discriminating line is shown in Figure 4.1, along with a subset of our galaxy sample that have SNR > 3 in each of the measured Hα, Hβ, [N II], and [O III] line fluxes, and no severe warning flags associated with these (we accept flags 0 and 1 in the scheme devised above). High values of log([N II]/Hα) and log([O III]/Hβ) are driven by a hard extreme ultraviolet (EUV) spectrum attributable to AGN activity in the host galaxies, pushing these galaxies to the top-right corner of the diagram. Although we can make an AGN classification on the basis of this diagram regardless of SNR, we use this technique only for SNR > 3, since below this, an increasing fraction of galaxies have measured line fluxes that are negative, and the non-symmetric distribution of galaxies on this diagram leads to classification biases. Typically, only a very small subset of our galaxies for which we performed spectral line measurements have all the required lines measured above our SNR criterion, but we are able to expand the classification by using only single line ratios. These are indicated by the dashed lines in Figure 4.1, and they correspond to log([N II]/Hα) > 0.3 and log([O III]/Hβ) > 1. Clearly these classifiers are less effective than that using both line ratios, however this does allow us to classify AGN according to their line ratios over a larger number of galaxies. We find only around 5% of our galaxies are classified as AGN, but we cannot rule out a small additional population that could not be identified in the manner just described.

After identifying AGN, we assume that the rest of the galaxies are regular star-forming or non star-forming galaxies. We identify star-forming galaxies as those that show measured fluxes that are positive and with SNR > 2 in any one of the Hα, Hβ, or [O II] lines. Those that do not meet this criterion are identified as non star-forming galaxies. Those galaxies that do not have good measurements of any of the Hα, Hβ, or [O II] emission lines (due to bad flags and/or lack of spectral coverage) are marked as 'unclassified'. These galaxies nevertheless have redshift measurements from e.g. the Ca II H and K, CH molecule G-band, Mg λ5175, and Na λ5894 lines.

For all other galaxies in our sample, we obtain their spectral classifications from the literature, typically found from principle component analyses (see Chen & Mulchaey, 2009; Prochaska et al., 2011a; Johnson et al., 2013; Tejos et al., 2014,

for details). Overall, we find that $\sim$ 55% of the galaxies in our sample are classified as star-forming, $\sim$ 35% are classified as non star-forming, $\sim$ 5% are classified as AGN, and $\sim$ 5% are unclassified.

Chapter 5

Simulated IGM absorber/galaxy samples

5.1 Overview

In this chapter, we present the methods used to extract a data set from a cosmological hydrodynamical simulation, which we then compare to the observational data. The comparison set comes from the Evolution and Assembly of GaLaxies and their Environments (EAGLE) project, which is in fact a suite of hydrodynamical simulations that follow the formation and evolution of galaxies and supermassive black holes in volumes representative of a ΛCDM Universe. In Section 5.2, we briefly describe the pertinent aspects of the simulation, and discuss its key advantages and limitations. We then follow in Section 5.3 with a detailed description of the processes involved in generating mock catalogues of galaxies and absorption systems, designed to mimic as closely as possible the observations.

5.2 The EAGLE simulations

The EAGLE project (Schaye et al., 2015) is a suite of cosmological hydrodynamical simulations representative of a ΛCDM Universe. The simulations were run with the smoothed particle hydrodynamics (SPH) code GADGET3 in cubic volumes 12.5, 25, 50, and 100 comoving Mpc on a side.[1] Numerous state-of-the-art numerical techniques and subgrid models are used to capture various physical processes important to galaxy formation and evolution. These include radiative gas cooling, star formation, mass loss from stars, metal enrichment, energy feedback from

[1]There are also a set of high resolution 'zoom' simulations (see Sawala et al., 2015, for details).

Table 5.1: Box sizes and resolutions of the main Eagle simulations, reproduced from Schaye et al. (2015).

Name	L (Mpc)[a]	N[b]	m_g ($M_\odot$)[c]	m_{dm} ($M_\odot$)[d]	ϵ_{com}[e]	ϵ_{prop}[f]
L025N0376	25	376^3	1.81×10^6	9.70×10^6	2.66	0.70
L025N0752	25	752^3	2.26×10^5	1.21×10^6	1.33	0.35
L050N0752	50	752^3	1.81×10^6	9.70×10^6	2.66	0.70
L100N1504	100	1504^3	1.81×10^6	9.70×10^6	2.66	0.70

[a] Comoving box size.
[b] Number of dark matter particles (there is an equal number of baryonic particles initially).
[c] Initial baryonic particle mass.
[d] Dark matter particle mass.
[e] Plummer-equivalent gravitational softening length.
[f] Maximum proper softening length.

supernovae and active galactic nuclei (AGN), and gas accretion onto, and mergers of, supermassive black holes. The efficiency of stellar feedback and the mass accretion onto black holes is calibrated to match the present-day stellar mass function of galaxies (subject to the additional constraint that the galaxies sizes need to be reasonable), and the efficiency of AGN feedback is calibrated to match the observed relation between stellar mass and black hole mass. Calibrations such as these are necessary, since the underlying physics behind galaxy feedback is neither well understood, nor well constrained observationally.

The box sizes, resolutions, and gravitational softening lengths[2] of the main Eagle simulations are listed in Table 5.1. Particle properties in each of the simulation volumes are recorded at 29 points in time between redshifts 20 and 0. These are referred to as 'snapshots'. In this thesis, we use only the largest simulation volume, L100N1504, and the snapshots between redshift 1 and 0. The resolution of Eagle is marginally sufficient to resolve the Jeans scales in the warm interstellar medium (ISM). The Jeans mass for a cloud with gas fraction f_g is

$$M_J \approx 10^7 f_g^{\frac{3}{2}} \left(\frac{n_H}{10^{-1}\ \mathrm{cm}^{-3}}\right)^{-\frac{1}{2}} \left(\frac{T}{10^4\ \mathrm{K}}\right)^{\frac{3}{2}} M_\odot, \tag{5.2.1}$$

[2]To prevent the occurrence of singularities in the force equation for N-body dynamics in close encounters between particles, the gravitational interaction is modified ('softened') on small scales according to some functional form. In Eagle, the functional form for this softening is similar to a Plummer profile.

and the Jeans length is

$$L_{\rm J} \approx 2 f_{\rm g}^{\frac{1}{2}} \left(\frac{n_{\rm H}}{10^{-1}\ {\rm cm}^{-3}} \right)^{-\frac{1}{2}} \left(\frac{T}{10^4\ {\rm K}} \right)^{\frac{1}{2}}\ {\rm kpc}. \tag{5.2.2}$$

If we take fiducial values of $f_{\rm g} \approx 1$, $n_{\rm H} \approx 0.5\ {\rm cm}^{-3}$, and $T \approx 5000$ K for the warm ISM, then we find $M_{\rm J} \approx 1.6 \times 10^6 M_{\odot}$ and $L_{\rm J} \approx 0.6$ kpc, which compare favourably with the gas particle masses and maximum proper gravitational softening lengths listed in Table 5.1. We shall refer to three particle types in Eagle: dark matter particles, star particles, and gas particles. Dark matter particles are evolved using the N-body part of the code, which simulates just the gravitational interactions between particles, while gas particles are also subject to hydrodynamical forces. Gas particles above a metallicity-dependent density threshold are converted to star particles stochastically, following the observed Kennicutt-Schmidt star-formation law (Kennicutt, 1998b). More details can be found in Schaye et al. (2015).

Note that there exists a high resolution variant of the 25 Mpc volume as part of the Eagle suite, which allows for an investigation into numerical convergence with resolution. For Eagle it is worth noting that there is only a demand for so-called 'weak convergence'. The meaning of this, and the motivation behind it, is discussed at length in Schaye et al. (2015) (see their section 2.2), but in brief this essentially means that simulation variants at differing resolution have the subgrid physics re-calibrated. Therefore, the expectation is that only the *re-calibrated* runs at higher resolutions to the nominal one should give converged predictions. This differs from the condition of so-called 'strong convergence', whereby numerical convergence must be achieved *without* the need for re-calibration.

Much like in other recent cosmological hydrodynamical simulations (e.g. Schaye et al., 2010; Vogelsberger et al., 2014), the key to producing realistic simulated galaxies in Eagle has mainly been the implementation of more effective subgrid models for stellar feedback, and the inclusion of subgrid models for AGN feedback (see Chapter 1, Section 1.5 for a discussion). Although much work has been devoted to improving the numerical techniques used to solve the hydrodynamical equations, it is variations on the implementation of these subgrid models that has

the largest impact on simulated galaxy properties (Scannapieco et al., 2012).

As already stated, the subgrid implementations of stellar and AGN feedback in EAGLE are calibrated against observed properties of the present-day galaxy population, as is standard practice. It is important to point out that this calibration offers fairly limited insight into the physics driving these feedback processes. Consider, for example, two simulations with the same treatment of the hydrodynamics, but run with differing resolution. Unfortunately, it often turns out that the best-calibrated subgrid models for feedback in each case can have substantially different efficiencies. In particular, the effectiveness of feedback is often hampered by excessive radiative losses. This problem is resolution dependent, to do with the numerical implementation, and depends on how realistic the modelling of the ISM is (e.g. Dalla Vecchia & Schaye, 2012). The inferred feedback efficiencies from calibration could be erroneous for other reasons too, for example if the energy/momentum in the subgrid model is coupled to the gas at the wrong scale, or if too much of the momentum that is injected cancels out. Therefore, even if the subgrid prescriptions for feedback produce a Universe that matches the observations perfectly, we cannot obtain constraints on the underlying physics governing those feedback processes.

Calibrating subgrid models for feedback has an important impact on the predictive power of cosmological hydrodynamical simulations with regards to galaxy properties as well. A key insight from these simulations is that there likely exists an equilibrium between the rate at which gas falls onto galaxies, the rate at which that gas is turned into stars, and the rate at which it is ejected. This is seen, for example, as driving the relation between stellar mass and gas-phase metallicity, the abundance of damped Lyman alpha absorbers, and the linear relation between star formation and H_2 abundance (e.g. Finlator & Davé, 2008; Schaye et al., 2010; Davé et al., 2012; Altay et al., 2013; Dekel et al., 2013; Feldmann, 2013; Lilly et al., 2013). In this picture, galaxy evolution is self-regulated - the gas outflow rate is determined by the rate of infall, not by the feedback efficiency. If gas outflows are driven by feedback from star formation, the star formation rate simply adjusts over time such that a balance between gas infall and outflow is achieved.

Similarly, if the outflows are driven by AGN feedback, then the black hole accretion rate adjusts over time to achieve this balance. Since star formation activity builds up the stellar mass in galaxies, and black hole accretion builds up the black hole mass, then it is *these* processes that are ultimately sensitive to the efficiency of the subgrid feedback implementation. This explains why subgrid feedback in EAGLE is calibrated to the galaxy stellar masses and black hole masses inferred from observations. Since there is a need for calibration, it is clear that a cosmological hydrodynamical simulation such as EAGLE cannot make predictions for these properties.

In view of the shortcomings outlined above, we are motivated to test the feedback prescriptions in EAGLE by examining its predictions for the distribution and dynamics of O VI absorbers around galaxies, which we can then compare to our observational sample. A test such as this has considerable diagnostic power, since the simulation was not calibrated to match observations such as these. Even though we can't hope to learn anything about the detailed physics governing supernova and AGN feedback, a simulation that matches these observations should nevertheless provide important insights on the gas flows around galaxies, which are responsible for driving the evolution of key galaxy properties, such as their star formation rates, and the mass-metallicity relationship. To perform this test, we need to generate mock catalogues of galaxies and absorbers from the simulation, which is the subject of the next section.

5.3 Creating the mock catalogues

In creating mock catalogues of galaxies and absorbers, we aim to mimic as closely as possible the observational procedure, and to achieve a sample whose selection biases are matched to those of the observational one. We do this to ensure a fair comparison between the real, and the simulated data. In addition, it is instructive to assess the basic statistical properties of the simulated samples. If these do not broadly match those observed, then this compromises our ability to draw any meaningful conclusions between a comparison of more complex diagnostics such

as their clustering amplitudes. In the following sections, we describe the data extraction procedures, and present some preliminary comparisons.

5.3.1 Galaxies

Galaxies in EAGLE are defined by the centre of the gravitational potential in 'subhaloes' identified by the SUBFIND algorithm (Springel et al., 2001; Dolag et al., 2009). A 'subhalo' is defined in this algorithm as a gravitationally bound, locally overdense group of particles, identified in the following fashion. First, dark matter haloes are found by implementing the friends-of-friends (FoF; Davis et al., 1985) algorithm on the dark matter particles. Any two particles less than a linking length of 0.2 times the mean inter-particle separation are called 'friends'. A dark matter halo is then a FoF group, defined by the set of particles for which each particle in the set is connected to every other particle via a network of friends. Gas and star particles in the simulation are then assigned to the same (if any) FoF halo as their nearest dark matter particles. Next, substructure candidates are identified by finding overdense regions within each FoF halo that are bounded by saddle points in the density distribution. Finally, particles not gravitationally bound to the substructure are discarded, and the resulting substructures are termed subhaloes. For EAGLE it was also necessary to merge subhaloes separated by less than the minimum of 3 kpc and the stellar half-mass radius, to remove a small number of very low mass subhaloes whose mass was dominated by a single particle (Schaye et al., 2015).

The stellar mass of each galaxy in EAGLE is defined to be the sum of the masses of all the star particles that belong to the corresponding subhalo, and that are within a 3D aperture of radius 30 kpc. The choice of aperture was motivated by the fact that this gives a nearly identical galaxy stellar mass function as do the Petrosian apertures used in the observations (Schaye et al., 2015). All galaxy properties in EAGLE, including the star formation rates, are evaluated in this aperture. For the purpose of matching to the observational sample, we limit the EAGLE sample to galaxies with stellar masses $\log(M_\star/M_\odot) > 8$. We perform this mass cut for two reasons: (i) to eliminate galaxies consisting of $\lesssim 100$ star particles, as these galax-

ies will have star-formation rates that are not statistically robust; and (ii) to reflect our observed sample, for which stellar masses $\log(M_\star/M_\odot) < 8$ are comparatively rare due to the limiting magnitudes of the surveys (see, for example, Li & White, 2009; Baldry et al., 2012).

To split the EAGLE sample of galaxies in terms of their star-formation activity, we adopt the same scheme as outlined in Schaye et al. (2015), whereby galaxies with specific star-formation rates $\dot{M}_\star/M_\star > 0.1$ Gyr^{-1} are defined to be star-forming galaxies, and all others are defined to be non-star-forming galaxies. We do not attempt to find a cut in star-formation rate (or specific star-formation rate) that will match the relative proportions of star-forming and non star-forming galaxies in the observational sample, since the observational definition is subject to signal-to-noise constraints not present in the simulation. We leave a more precise comparison for future work.

Finally, we define the z-axis of the simulation volume to be the pseudo-redshift axis, and replace the z-coordinate of each galaxy with $z_{\rm pec} = z + v_{\rm pec}(1+z)/H(z)$, where $v_{\rm pec}$ is the peculiar velocity of the galaxy, and $H(z)$ the Hubble parameter for a redshift z corresponding to the snapshot redshift. This then mimics the distortions introduced by peculiar velocities along the line-of-sight (LOS) for galaxy surveys constructed in redshift space.

5.3.2 Absorption-line systems

Observationally, the diffuse baryonic material in the Universe - the intergalactic medium (IGM) - is characterised by measurements of absorption-line systems in quasar (QSO) spectra (see Chapter 2). We therefore characterise the IGM in EAGLE in a very similar manner, by drawing synthetic QSO sight-lines through the simulation volume. We follow the procedure outlined in Theuns et al. (1998) (see their appendix A4), using a modified version of the artificial transmission spectra code, SPECWIZARD. This works as follows. For a given (x, y, z) coordinate and orientation in the simulation volume at a given redshift snapshot, specifying a one-dimensional sight-line, SPECWIZARD first extracts all SPH gas particles that intersect that sight-line. The sight-line is then divided into velocity bins labelled

from zero to $\dot{a}L$, where $a(z)$ is the dimensionless scale-factor as a function of redshift, and L is the box size in comoving coordinates. Each bin has a velocity width $\Delta = 1$ km s^{-1}. For each bin, the code then calculates the local physical density, ρ_X, and temperature, T_X, for an ionic species X, weighted by the SPH smoothing kernel and abundance of species X, assuming ionization equilibrium in the presence of a Haardt & Madau (2001) ultraviolet (UV) background radiation field. Then, for a given atomic transition, i, of species X, assuming only thermal line broadening, by analogy with equations (1.4.42) and (1.4.47), a bin k, corresponding to a velocity $v(k)$, will suffer absorption due to material in bin j, at velocity $v(j)$, by an amount $\mathrm{e}^{-\tau(k)}$, where

$$\tau(k) = \sigma_{X_i} \frac{1}{\sqrt{\pi}} \frac{c}{V_X(j)} \rho_X(j) a \Delta \exp\left[-\left(\frac{v(k) - v(j)}{V_X(j)} \right)^2 \right], \tag{5.3.3}$$

and

$$V_X(j) = \frac{2kT_X(j)}{m_X} \tag{5.3.4}$$

(Theuns et al., 1998). Here, σ_{X_i} is the absorption cross-section of the transition, c is the speed of light, and $V_X(j)$ is the Doppler width of species X with mass m_X. For the vast majority of the absorption along these sight-lines, the physical densities are small enough that a purely thermally broadened line-profile is a good approximation to the real one.

To create mock catalogues of O VI absorbers, we use the method above to calculate the optical depth, $\tau(v)$, in O VI λ1031 along 25 000 randomly-drawn sight-lines parallel to the z-axis (our pseudo-redshift axis) through each of 7 different redshift snapshots of the 100 Mpc volume over the range $0.1 \lesssim z \lesssim 0.7$ (the dominant range covered by our observational sample). We then take peaks in the τ distribution above a threshold value of 0.0005 (arbitrary) along each sight-line to correspond to the optical depth at the absorption line centres, and calculate the absorbing column density using equation (1.4.50), assuming a Doppler broadening parameter equivalent to V_X in the equations above. Each absorber is then

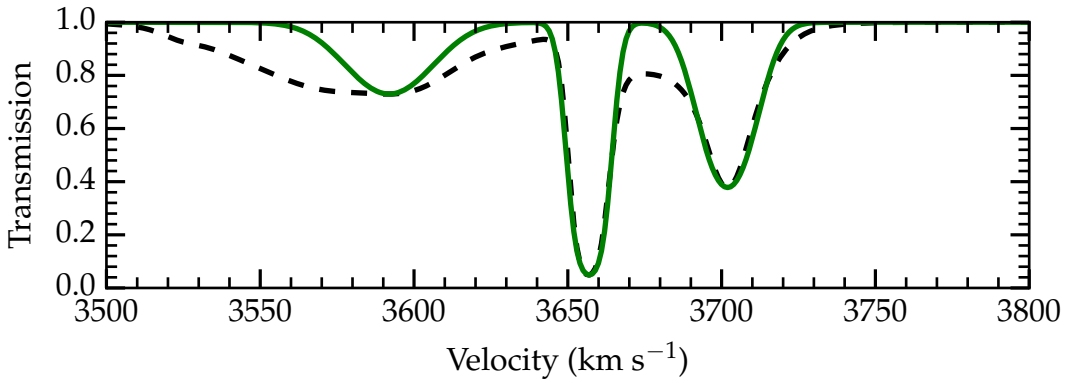

Figure 5.1: Voigt profiles derived from peaks in the optical depth distribution (plotted in green) compared to the $e^{-\tau}$ spectrum. The predicted Voigt profiles show simpler velocity structure than is apparent in the spectrum, highlighting an important caveat to our approach.

assigned the (x, y) coordinate of the sight-line it was extracted from, and the velocity, v, at which it was extracted, which we convert to a position r_z along the z-axis via $r_z = v(1 + z)/H(z)$, where z is the redshift of the simulation snapshot, and $H(z)$ is the value of the Hubble parameter at that redshift.

It is important to note that the above procedure differs from the observational one, whereby Voigt profiles are fit to the absorption features in transmission spectra ($e^{-\tau}$) in order to extract column densities and Doppler broadening parameters. This approach enables de-blending of multiple-component absorption, and takes into account the broadening of the lines due to the instrumental profile, enabling accurate recovery of their column densities. The technique was originally devised under the premise that the intervening absorption lines in QSO spectra arose from discrete absorbing clouds, but this picture was challenged early by the smoothly distributed IGM captured in cosmological hydrodynamical simulations (e.g. Cen et al., 1994; Hernquist et al., 1996; Miralda-Escudé et al., 1996; Theuns et al., 1998; Davé et al., 1999). Absorbers that have a large spatial extent take part in the Hubble expansion, which leads to a line profile that deviates from a Voigt profile. Voigt profile fitting the transmission spectra will therefore glean slightly different results to simply taking peaks in the τ distribution. In particular, we might expect a slightly larger number of absorbers, and some differences in the derived Doppler

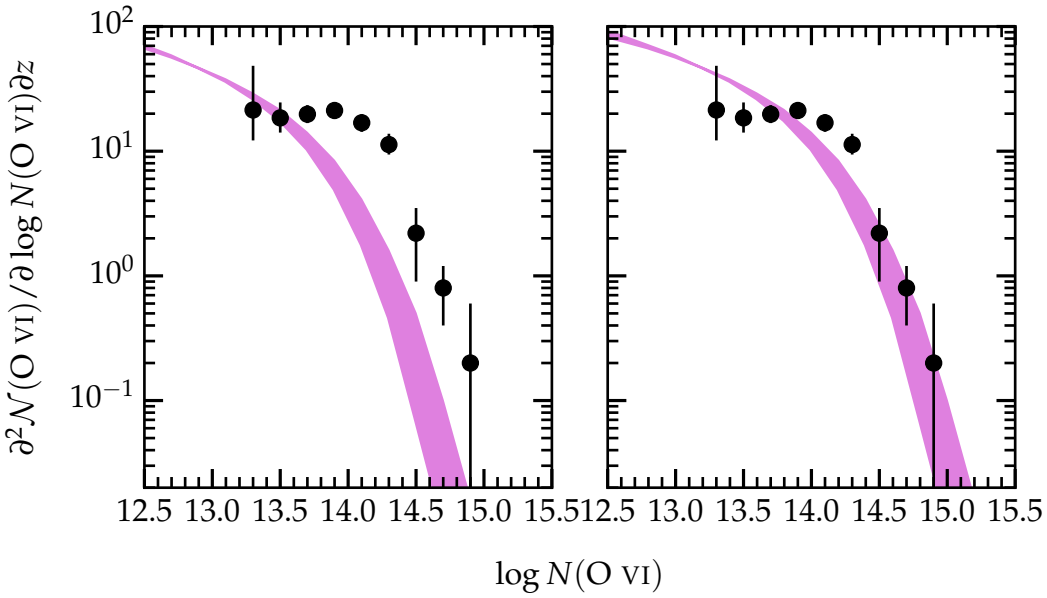

Figure 5.2: The column density distribution function of O VI absorbers from Danforth et al. (2014) (data points) compared to predictions from EAGLE snapshots over the redshift range $0.1 \lesssim z \lesssim 0.7$ (shaded magenta region). The shape of the column density distribution function predicted from EAGLE is not far from that in the data (left panel), but a shift of 0.3 dex in column density (right panel) is necessary to obtain the required number of high column density absorbers from the simulation (see the text for justification).

broadening parameters and column densities. As a check, we derived Voigt profiles for lines recovered in some of the sight-lines extracted from the simulation using our τ peak method, and plotted these on top of the transmission spectrum. An example is shown in Figure 5.1. The transmission spectrum is shown as the black dashed line, and the predicted Voigt profiles are shown in green. More structure is apparent in the real spectrum, which would yield a larger number of Voigt components in Voigt profile fitting, and column densities and Doppler broadening parameters that differ slightly from those we have recovered via the τ peak method. This an obvious caveat to our approach, which we bear in mind when interpreting our results later on.

Having created a mock catalogue of absorbers from the EAGLE simulation, we now examine whether the global statistics of our mock population matches the observed one. To do so, we investigate the column density distribution function of O VI absorbers. This is defined as the number, $\mathcal{N}$, of absorption lines per unit

column density, $\mathrm{d}N_{\mathrm{O\,VI}}$, per unit redshift, $\mathrm{d}z$:

$$f(N_{\mathrm{O\,VI}}) = \frac{\mathrm{d}^2\mathcal{N}}{\mathrm{d}N_{\mathrm{O\,VI}}\mathrm{d}z}. \tag{5.3.5}$$

It is also sometimes defined in terms of the absorption distance $\mathrm{d}X$, which is related to $\mathrm{d}z$ via $\mathrm{d}X/\mathrm{d}z = H_0(1+z)^2/H(z)$, where $H(z)$ is the Hubble parameter. We compare to the measurement presented in Danforth et al. (2014) (data points in Figure 5.2), whose data form the majority of our O VI sample. To calculate this for the O VI absorbers extracted from EAGLE, we adopt the same bin size as was used for the data, $\Delta(\log N_{\mathrm{O\,VI}}) = 0.2$, and note that we have $n_{\mathrm{LOS}} = 25\,000$ sight-lines through the $L = 100$ Mpc simulation volume, then calculate $\Delta z = n_{\mathrm{LOS}}LH(z)/c$, where c is the speed of light. We do this for each of the redshift snapshots in the range $0.1 \lesssim z \lesssim 0.7$, and plot the resulting curves as the shaded region in the left-hand panel of Figure 5.2. The shape of the column density distribution function from EAGLE is not far from that of the observed one, although the 'knee' seen in the observed distribution is not so well pronounced in the simulation. We note also that the distribution of simulated O VI absorbers falls off at lower column densities than in the data. Uncertainties in the shape and normalisation of the UV background are likely enough to account for this difference, along with uncertainties in the metal yields from star-formation (Schaye et al., 2015). Thus, given the broad agreement in the shape of the column density distribution function between the simulation and the data, we are motivated to shift the simulated O VI absorption column densities by 0.3 dex towards higher values, resulting in a greater concordance with the data (right-hand panel in Figure 5.2). This shift is also necessary if we are to draw a sub-sample of O VI absorbers from the simulation that is matched to the selection functions in the data. We therefore proceed from this point on with simulated O VI absorbers whose column densities are rescaled by an extra 0.3 dex in column density.

To create a matched sample of O VI absorption systems from the EAGLE simulation, we begin by optimally binning the column density histogram of the observed sample using the method presented in Knuth (2006) (see Section 6.4 in Chapter 6

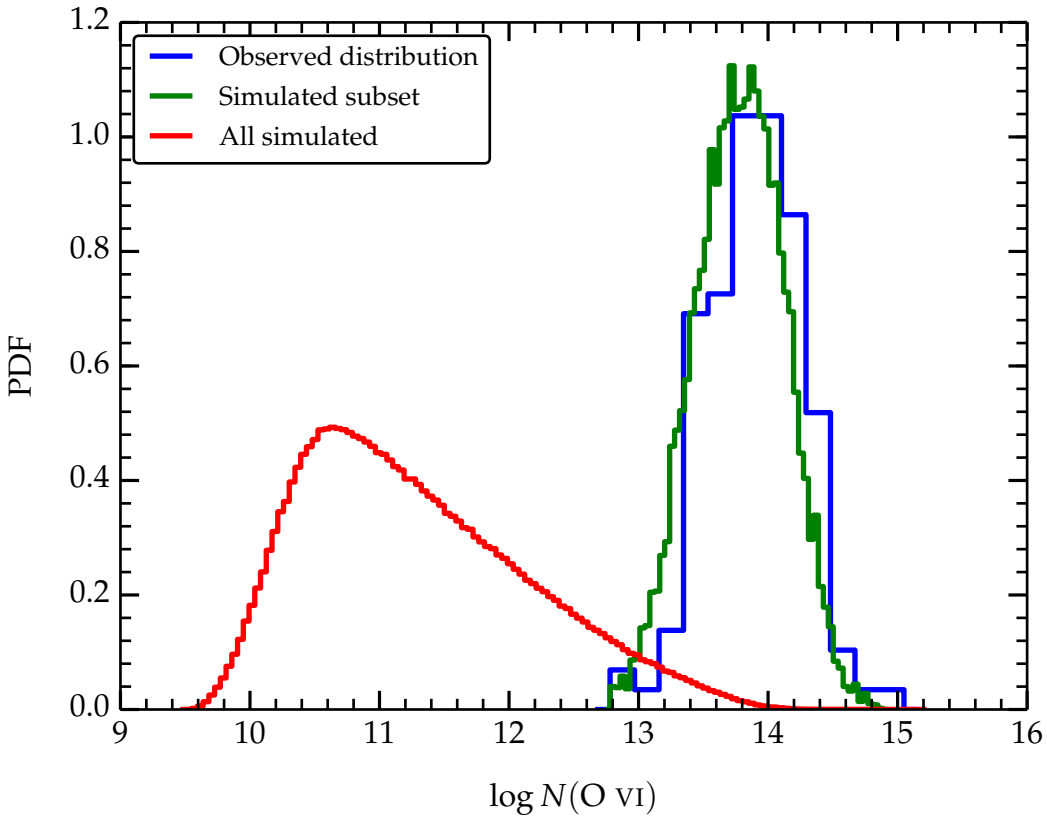

Figure 5.3: Probability density functions for the observed O VI column densities (blue), the simulated ones (red) and the simulated subset (green). See the text for a full description.

for a description and motivation), then smooth the histogram with a Gaussian having standard deviation equal to the bin size. We then interpolate a cubic spline over the smoothed histogram, and normalise by the area underneath the curve to produce a probability density function (PDF) for the observed column densities. We also do the same on the column density distribution of simulated samples, which exhibits a turn over at very low column densities ($N(\mathrm{O\ VI}) < 10^{11}$ cm^{-2}) due to the arbitrarily imposed optical depth cut. We then randomly draw a 10% subset of the simulated O VI absorbers from the probability density distribution of the data, inverse weighted by that of the simulation. We illustrate this in Figure 5.3. The observed probability density function is shown in blue, the simulated one in red, and that of the 10% subset in green. As is clearly evident from the figure, we are able to extract a sub-sample of O VI absorbers from the EAGLE simulation that reproduces well the selection bias towards high column densities in the data.

Chapter 6

IGM-galaxy clustering at $z < 1$

6.1 Overview

In this chapter, we present new results on the auto- and cross-correlation functions of galaxies and O VI absorbers in a ~ 18 Gpc^3 comoving volume at $z < 1$. We use a sample of 51 296 galaxies and 140 O VI absorbers to measure two-point correlation functions in the two dimensions transverse and orthogonal to the line-of-sight $\xi(r_\perp, r_\parallel)$. We furthermore infer the corresponding 'real-space' correlation functions, $\xi(r)$, by projecting $\xi(r_\perp, r_\parallel)$ along $r_\parallel$, and assuming a power-law form, $\xi(r) = (r/r_0)^{-\gamma}$. Comparing the results from the absorber-galaxy cross-correlation function, $\xi_{\rm ag}$, the galaxy auto-correlation function, $\xi_{\rm gg}$, and the absorber auto-correlation function, $\xi_{\rm aa}$, we constrain the statistical connection between galaxies and the metal-enriched intergalactic medium as a function of star-formation activity. We also compare these results to predictions from the Evolution and Assembly of GaLaxies and their Environments (EAGLE) cosmological hydrodynamical simulation. Our results are consistent with the following conclusions: (i) O VI absorbers show very little velocity dispersion with respect to galaxies on $\sim$ Mpc scales, likely $\lesssim 100$ km s^{-1}; (ii) O VI absorbers and galaxies may not linearly trace the same underlying distribution of matter in general. In particular, the distribution of O VI around galaxies could be more extended than the distribution of galaxies around themselves; (iii) O VI absorbers are equally likely to inhabit the same regions as star-forming galaxies as they are non star-forming galaxies; and (iv) O VI absorbers are either not ubiquitous to galaxies, or their distribution around them is patchy on scales $\gtrsim 100$ kpc (or both), at least for the column densities at which most are currently detected. We find a striking agreement between

the observational results and those predicted from the EAGLE simulation. Our results therefore suggest that cosmological hydrodynamical simulations are able to make robust predictions for the distribution and dynamics of metal-enriched gas around galaxies.

6.2 Introduction

The connection between the intergalactic medium (IGM) and galaxies is fundamental to our understanding of the formation and evolution of galaxies and the large-scale structure of the Universe. This is because there exists a continuous interplay between galaxies and the plasma around them, which fuels the formation of stars and the hierarchical assembly of cosmic structures. Under a theoretical Λ cold dark matter paradigm, the two main physical processes that drive this assembly are: (i) the accretion of intergalactic matter in 'hot' and 'cold' modes (e.g. Rees & Ostriker, 1977; White & Rees, 1978; White & Frenk, 1991; Kereš et al., 2005; van de Voort et al., 2011); and (ii) winds emanating from galaxies generated mostly by supernova (SN) explosions and active galactic nuclei (AGN) (e.g. Baugh et al., 2005; Veilleux et al., 2005; Bower et al., 2006; Lagos et al., 2008; Creasey et al., 2013). These winds are also thought to be responsible for enriching the IGM with metals. Observational studies are producing results largely consistent with this picture, but do not yet conclusively demonstrate it to be true.

To achieve a detailed understanding on the formation of galaxies and cosmic structure, we must correctly describe the behaviour and evolution of the baryonic matter in the Universe. For this we require hydrodynamical simulations following the evolution of baryons and dark matter together within a cosmological volume (e.g. Crain et al., 2009; Davé et al., 2010; Vogelsberger et al., 2014; Schaye et al., 2015). Large volumes are fundamentally important, since the simulations must be able to reproduce the statistics of the present day galaxy population. Unfortunately, due to the computational cost, there is a fundamental reliance on uncertain 'sub-grid' prescriptions to capture the relevant physics on scales smaller than the resolution limit (e.g. Schaye et al., 2010; Scannapieco et al., 2012, see also

Chapter 1, Section 1.5, and Chapter 5, Section 5.2). To glean trust-worthy physical insight from these simulations, we must therefore place tight constraints on the sub-grid physics that must inevitably be included. Observations of the gaseous environments around galaxies play a major role in this goal, as the simulations are not typically calibrated to match these observations. They therefore provide an important test that is independent of any 'fine tuning'.

Unfortunately, despite being the main reservoir of baryons at all epochs, the extremely low density of the IGM makes its observation both limited and difficult. The only viable method at present is through the analysis of absorption lines in quasar (QSO) spectra (see Chapter 2). These lines appear due to the scattering of ultraviolet (UV) photons by intervening gas along the line-of-sight (LOS) (see Chapter 1, Section 1.4.2). The resulting characterisation of the IGM is therefore limited to being one-dimensional. Nevertheless, by combining information from multiple LOS, we are able to construct a statistical picture of the distribution and dynamics of gas in the Universe.

Observations of the IGM at low redshifts ($z < 1$) have opened up dramatically over the last few years with the advent of the Cosmic Origins Spectrograph (COS) on the Hubble Space Telescope (*HST*) (Green et al., 2012, see also Chapter 2, Section 2.2). With a sensitivity more than 10 times that of its predecessor, COS has provided obervations of hundreds of QSOs in the far ultraviolet (FUV). Observations at these wavelengths are fundamentally important, as they allow for a mapping of the H I and metal content of the IGM to $z = 0$. The capabilities of COS have been exploited extensively to probe both cool ($T \sim 10^4$ K) gas, traced mostly by the Lyα forest, and warmer ($\sim 10^5 - 10^6$ K) gas, traced by broad Lyα, O VI, and Ne VIII absorption. These ions probably trace up to $\sim 60\%$ of all the baryons, with only $\sim 10\%$ in the luminous constituents of the Universe (stars and galaxies), and the rest in an even hotter plasma at $T > 10^6$ K.

Much of the work on the low redshift IGM in relation to galaxies to date has taken a 'galaxy-centric' approach, with a primary focus on the properties of the so-called circumgalactic medium (CGM). A number of successful programs have been designed with this goal in mind, notably the 'COS-Halos' survey (Tumlin-

son et al., 2013), and various programs by the COS guaranteed time observations (GTO) team (e.g. Stocke et al., 2013; Keeney et al., 2013). These studies implicitly assume a one-to-one correspondence between absorption systems and the closest observed galaxy, which is problematic due to the incomplete sampling of galaxies in any galaxy survey. Despite this shortcoming, it is clear from these studies that there is a nearly ubiquitous presence of cool ($T \approx 10^4 - 10^5$ K) metal-enriched gas surrounding galaxies to impact parameters of ~ 150 kpc (see e.g. Prochaska et al., 2011b; Werk et al., 2013). Ionization models suggest this cool CGM, combined with an additional hotter component traced by O VI, can account for at least half of the baryons expected from Big Bang nucleosynthesis that were originally unaccounted for (Fukugita et al., 1998; McGaugh et al., 2010; Werk et al., 2014). Nevertheless, 30 – 40% of the baryons may still be unaccounted for, residing in the so-called warm-hot intergalactic medium (WHIM) predicted by cosmological hydrodynamical simulations (e.g. Cen & Ostriker, 1999; Davé et al., 2001). An unambiguous detection of the WHIM is needed if we are to validate the predictions of these simulations.

In this chapter, we shall address the connection between the metal-enriched IGM and galaxies at $z < 1$ via an analysis of their two-point cross- and auto-correlation functions. We shall use the O VI $\lambda\lambda 1031, 1037$ doublet to trace the metal-enriched components of the IGM, as these are the most commonly observed metal transitions at low redshifts. They are also known to trace a wide range of gas phases and environments (see Chapter 2, Section 2.8 for a discussion). Using these measurements, we shall investigate the distribution and dynamics of the metal-enriched IGM around galaxies on both the CGM scale ($\lesssim 300$ kpc) and to much larger scales ($\gg 1$ Mpc). In addition, we shall critically compare our observational measurements to predictions from the Eagle cosmological hydrodynamical simulation.

6.2.1 Outline

We shall make use of the sample of O VI absorbers described in Chapter 2, the galaxy surveys described in Chapter 4, and the data sets drawn from the Eagle

cosmological hydrodynamical simulation, described in Chapter 5, to examine the two-point correlation functions of O VI absorbers and galaxies. We limit the QSO sight-line fields listed in Table 4.2 to just those where we found O VI absorption in at least one QSO spectrum with confidence label 'a' or 'b' (see Chapter 2, Section 2.8). We also exclude O VI absorbers within 5000 km s^{-1} of the QSO redshift, on the conservative assessment that these may be intrinsic to the QSO host galaxy. Galaxies are limited to those with redshift confidence labels 'a' or 'b' (see Section 4.3). The resulting sample for this study then consists of 27 fields and 32 QSO sight-lines, which contain 51 296 galaxies and 140 O VI absorbers respectively. Our full sample of galaxies and QSOs will be used to investigate the H I-galaxy cross-correlation function in future work. This will serve as an update to the work presented in Tejos et al. (2014).

In Section 6.3, we describe the mathematical formalisms used to measure the O VI-galaxy cross-correlation and their respective auto-correlation functions. In Section 6.4, we describe the process of generating random samples from our data. In Section 6.5, we present our results, and in Section 6.6, we discuss the implications of our findings, and draw conclusions.

6.3 Correlation functions

To address the statistical connection between the metal-enriched IGM (traced by O VI absorbers) and galaxies, we focus on a two-point correlation analysis. The advantages of doing so are two-fold: (i) we do not rely on associating a particular intergalactic absorber with a particular galaxy (or set of galaxies), which in many instances is ambiguous; and (ii) we are robust to galaxy/absorber completeness variations in our survey, since we are measuring a clustering excess as a function of scale relative to a random expectation that takes into account the survey selection function. For the latter, it is important to note that a deficit of one population relative to another at a given scale due to a selection bias *will* bias the measurement at those scales if it is not properly modelled and corrected for.

The two-point correlation function, $\xi(r)$, is defined as the probability *excess* of

finding two points separated by a distance r with respect to the random expectation.[1] We shall use the correlation function between pairs of galaxies, $\xi_{\rm gg}$ (the galaxy auto-correlation function), between pairs of absorbers, $\xi_{\rm aa}$ (the absorber auto-correlation function) and between absorber-galaxy pairs, $\xi_{\rm ag}$ (the absorber-galaxy cross-correlation function) to gain insights on the relationship between them.

For each field, we assign a central coordinate in right-ascension and declination (α_0, δ_0). Then, for every object in that field with a spectroscopic redshift z, we calculate its position in Cartesian (x, y, z) coordinates as follows:

$$\begin{aligned} x &\equiv r(z)\cos(\Delta\delta)\cos(\Delta\alpha), \\ y &\equiv r(z)\cos(\Delta\delta)\sin(\Delta\alpha), \\ z &\equiv r(z)\sin(\Delta\delta), \end{aligned} \tag{6.3.1}$$

where $r(z)$ is the comoving distance, as defined in equation (1.2.21), $\Delta\delta \equiv (\delta - \delta_0)$ and $\Delta\alpha \equiv (\alpha - \alpha_0)\cos(\delta_0)$, both in radians. All of our fields are away from the celestial poles and have small angular coverage, making this transformation accurate. The x coordinate is parallel to the LOS, while the y and z coordinates are perpendicular (transverse) to it, so given that peculiar velocities contribute to the redshifts of objects in addition to cosmological expansion, our coordinate system is subject to distortions, often termed 'redshift-space distortions' (e.g. Kaiser, 1987). We therefore measure correlation functions parallel and transverse to the LOS independently, i.e. we measure the two-dimensional two-point correlation function $\xi(r_\perp, r_\parallel)$, where for a given pair of objects denoted i and j, we have

$$\begin{aligned} r_\perp &\equiv \sqrt{|y_i - y_j|^2 + |z_i - z_j|^2}, \\ r_\parallel &\equiv |x_i - x_j|. \end{aligned} \tag{6.3.2}$$

Deviations from an isotropic signal in these coordinates can then be attributed to peculiar velocities along the LOS and/or large-scale bulk motions between ob-

[1] Assumes isotropy, since r in this case is a scalar quantity.

jects in our sample. Nevertheless, it is also instructive to measure the correlation function for three-dimensional separations in redshift space, $\xi(s)$, where

$$s = \sqrt{r_\perp^2 + r_\parallel^2}. \tag{6.3.3}$$

We use the Landy & Szalay (1993) (LS) estimator to calculate the auto-correlation functions:

$$\xi(r_\perp, r_\parallel) = \frac{DD/n_{\rm DD} - 2DR/n_{\rm DR}}{RR/n_{\rm RR}} - 1, \tag{6.3.4}$$

where DD, DR, and RR are the number of data-data, data-random and random-random pairs respectively at a given $r_\perp$ and $r_\parallel$ (or at a given s for the redshift-space correlation function), and the values of n correspond to their normalisation factors:

$$\begin{aligned} n_{\rm DD} &= N(N-1)/2, \\ n_{\rm DR} &= \vartheta N^2, \\ n_{\rm RR} &= \vartheta N(\vartheta N - 1)/2, \end{aligned} \tag{6.3.5}$$

where N is the total number of real objects, and ϑN is the total number of random ones. We write the normalisation factors in this way to highlight that the random samples always have an integer number ϑ times as many objects as the real ones. For the absorber-galaxy cross-correlation function, we adopt the following, generalised form of the LS estimator:

$$\xi_{\rm ag}(r_\perp, r_\parallel) = \frac{D_{\rm a}D_{\rm g}/n_{\rm ag}^{\rm DD} - D_{\rm a}R_{\rm g}/n_{\rm ag}^{\rm DR} - R_{\rm a}D_{\rm g}/n_{\rm ag}^{\rm RD}}{R_{\rm a}R_{\rm g}/n_{\rm ag}^{\rm RR}} - 1 \tag{6.3.6}$$

(e.g. Adelberger et al., 2003; Tejos et al., 2014), where $D_{\rm a}D_{\rm g}$, $D_{\rm a}R_{\rm g}$, $R_{\rm a}D_{\rm g}$, and $R_{\rm a}R_{\rm g}$ are the data-data, data-random, random-data, and random-random absorber-galaxy

pairs respectively, and their normalisation factors, n, are

$$\begin{aligned} n_{\rm ag}^{\rm DD} &= N_{\rm a} N_{\rm g}, \\ n_{\rm ag}^{\rm DR} &= \vartheta_{\rm g} N_{\rm a} N_{\rm g}, \\ n_{\rm ag}^{\rm RD} &= \vartheta_{\rm a} N_{\rm a} N_{\rm g}, \\ n_{\rm ag}^{\rm RR} &= \vartheta_{\rm a} \vartheta_{\rm g} N_{\rm a} N_{\rm g}, \end{aligned} \tag{6.3.7}$$

where $N_{\rm a}$ and $N_{\rm g}$ are the total number of absorbers and galaxies respectively, and $\vartheta_{\rm a} N_{\rm a}$ and $\vartheta_{\rm g} N_{\rm g}$ are the total number of random absorbers and random galaxies respectively. Landy & Szalay (1993) have shown that these estimators minimise the variance in the correlation function, and so are preferable to other proposed estimators.

A useful quantity, which we also compute, is the projection of the two-dimensional two-point correlation function along the LOS:

$$\Xi(r_\perp) = 2 \int_0^\infty {\rm d}r_\parallel \; \xi(r_\perp, r_\parallel). \tag{6.3.8}$$

In reality, one only integrates $\xi(r_\perp, r_\parallel)$ up to a finite $r_\parallel$ where the correlation function ceases to be well measured, or where it is consistent with zero. The advantage of calculating this quantity is that it integrates over correlations smeared along the LOS due to peculiar motions, and thus it is insensitive to redshift-space distortions on small transverse scales where bulk flows are not important (Davis & Peebles, 1983). We can therefore find a relation between the 'real-space' correlation function (free of distortions), $\xi(r)$, and $\Xi(r_\perp)$ as:

$$\begin{aligned} \Xi(r_\perp) &= 2 \int_0^\infty {\rm d}r_\parallel \; \xi(r) \\ &= 2 \int_{r_\perp}^\infty {\rm d}r \; \xi(r) \frac{r}{\sqrt{r^2 - r_\perp^2}}, \end{aligned} \tag{6.3.9}$$

which gives $\xi(r)$ as the inverse Abel transform

$$\xi(r) = -\frac{1}{\pi}\int_r^\infty \frac{\mathrm{d}\Xi(r_\perp)}{\mathrm{d}r_\perp}\frac{\mathrm{d}r_\perp}{\sqrt{r_\perp^2 - r^2}}. \tag{6.3.10}$$

Davis & Peebles (1983) showed that when $\xi(r)$ is described by a power law of the form

$$\xi(r) = \left(\frac{r}{r_0}\right)^{-\gamma}, \tag{6.3.11}$$

equation (6.3.9) yields to

$$\Xi(r_\perp) = A(r_0, \gamma) r_\perp^{1-\gamma}, \tag{6.3.12}$$

where $A(r_0, \gamma) = r_0^\gamma \Gamma(1/2)\Gamma[(\gamma-1)/2]/\Gamma(\gamma/2)$, and Γ is the Gamma function. Fitting a power law form to $\Xi(r_\perp)$ therefore allows determination of r_0 and γ, and hence $\xi(r)$, for $\gamma > 1$. Here r_0 is usually referred to as the 'correlation length', and γ is the slope of the correlation function.

Given the volume-limited nature of any survey, all estimators are biased towards correlation amplitudes that are lower than the real ones. This arises because the mean density of objects is estimated from the survey itself, and is a well-known bias commonly referred to as the 'integral constraint' (Groth & Peebles, 1977). Landy & Szalay (1993) showed that the ξ measured using equation (6.3.4) or equation (6.3.6) and the real one, $\xi_{\rm real}$, are related as

$$1 + \xi = \frac{1 + \xi_{\rm real}}{1 + \xi_V}, \tag{6.3.13}$$

where ξ_V is the (scalar) integral constraint, defined as

$$\xi_V \equiv \int_V \mathrm{d}^2V\ G(r)\xi_{\rm real}(r) \tag{6.3.14}$$

Here, $G(r)$ is the normalised geometric window function, which gives the probability of having two volume elements separated by a distance r for a given survey

geometry. For a large enough random catalogue, this is accurately approximated as $G(r) \approx \mathrm{RR}/n_{\mathrm{RR}}$.

Although we cannot know ξ_{real} a priori, we can still estimate the integral constraint by obtaining r_0 and γ from a power-law fit to $\Xi(r_\perp)$, and taking the mean of the random-random pair counts over all $r_\parallel$ bins at each $r_\perp$ bin to give a proxy for ξ_V as

$$\tilde{\xi}_V = \sum_{r_\perp} \langle \mathrm{RR}/n_{\mathrm{RR}} \rangle A(r_0, \gamma) r_\perp^{1-\gamma}. \tag{6.3.15}$$

We can then make a (small) correction to our measured correlation function, ξ', to obtain ξ as follows:

$$\xi = \xi' + \tilde{\xi}_V(1 + \xi'). \tag{6.3.16}$$

All of the correlation function measurements that follow have this correction applied.

To interpret our correlation function measurements, we follow Adelberger et al. (2003) and Tejos et al. (2014) in using the Cauchy-Schwarz inequality:

$$\xi_{\mathrm{ag}}^2 \leq \xi_{\mathrm{gg}} \xi_{\mathrm{aa}}. \tag{6.3.17}$$

The equality can only hold at any given scale when the density fluctuations that give rise to absorbers and galaxies are linearly dependent. In other words, both populations must trace the same underlying distribution of matter with a linear bias (independent of the scale) to achieve $\xi_{\mathrm{ag}}^2 = \xi_{\mathrm{gg}} \xi_{\mathrm{aa}}$ in general.

We estimate the uncertainty in our correlation function measurements using the bootstrap method, which in our experience provides the most conservative measure of the uncertainty (see Tejos et al., 2014, for a discussion). We do this by creating $N_{\mathrm{bs}} = 1000$ sets of 27 fields, randomly chosen (with replacement) from

our set of 27 fields, and compute the uncertainty as

$$\Delta^2(\xi) = \frac{1}{N_{\rm bs}} \sum_{i}^{N_{\rm bs}} (\xi_i - \bar{\xi})^2, \tag{6.3.18}$$

where ξ_i is the correlation function measured from the ith random set of fields, and $\bar{\xi}$ is the mean of these measurements. Performing bootstrap realisations over fields ensures that we capture the sample variance, as well as the statistical uncertainty in the measurement.

The bootstrap uncertainty estimation is clearly appropriate for our observational sample, for which we have a large number of independent fields, but is not so easily applied to the simulated sample we have assembled, which is drawn from a single cubic volume, 100 comoving Mpc on a side (see Chapter 5 for details). To apply this uncertainty estimator, we would have to break the simulation down into sub-volumes, which limits the scales on which we can measure the correlation functions. For the simulated samples, we therefore quantify the statistical uncertainty on the measurement via the approximate estimator presented in Landy & Szalay (1993):

$$\Delta^2_{\rm LS}(\xi) \approx \frac{(1+\xi)^2}{n_{\rm DD}(RR/n_{\rm RR})} \approx \frac{(1+\xi)^3}{DD}. \tag{6.3.19}$$

Note that this greater than the commonly used Poissonian estimator, $\Delta^2_{\rm DD}(\xi) = (1+\xi)/{\rm DD}$, by a factor of $\sim (1+\xi)^2$, since it takes into account correlations introduced by non-independent cross-pairs. In reality, we are able to achieve negligibly small statistical uncertainties on the measurements from the simulated data, but we note that the sample variance (which may be much larger than the statistical one) is not taken into account.

6.4 Random catalogues

As is clear from the previous section, the construction of random samples is a crucial part of any correlation function analysis. These random samples need to

capture the selection functions in the data, so as not to bias the measurement. We present in the following sections a detailed description of the method for generating random samples of observed and simulated O VI absorbers and galaxies.

6.4.1 Random galaxy catalogues

For our observational sample, we create random galaxies for each field and survey independently. This means that individual fields containing galaxies from multiple surveys have separate random catalogues constructed for each of those surveys that are then combined at the end to form the random sample for that field. We do this since the different surveys in our sample have different selection functions (see Chapter 4, Table 4.1), and it is easier to model these separately, rather than attempting to model the combined selection functions. We note that the T14-Q0107 survey listed in Table 4.1 has a complex selection function, since it combines 4 sub-surveys constructed with different instruments, each with their own specific selection biases. Ideally, we would further split this survey down into its constituent parts, but the number of galaxies attributable to each of the different instruments is too small to reliably model their individual selection functions.

Our process for creating random galaxy catalogues for the observational sample expands upon that described in Tejos et al. (2014), and works as follows. For a given galaxy, in a given field, and from a given survey, we create $\vartheta_g = 10$ random ones, varying the redshift of the galaxy, but preserving its position on the sky and all of its other properties. The random redshifts are drawn from a probability density function that is modelled on the observed redshift histogram for galaxies with matching properties in the survey from which the real redshift was obtained. We take into account the observed magnitude of the galaxy (in the broadband filter that defines the magnitude limit for the survey), and whether it is a star forming galaxy, a non star-forming galaxy, or neither of these. For example, if a galaxy in a particular field is drawn from the Sloan Digital Sky Survey (SDSS), and is a star-forming galaxy with an r-band magnitude $r = 16$, we randomly draw $\vartheta_g = 10$ redshifts from a probability density function that is modelled on the redshift histogram in SDSS for star-forming galaxies with that magnitude. In this way, our

random samples reflect the survey sensitivity functions for galaxies of a given magnitude and spectral type, and also the evolution in the star-forming fraction. For galaxies that have no classification of spectral type, or are classified as AGN, we use the redshift histogram for *all* galaxies with the same magnitude as that galaxy. Our approach guarantees that we take into account the survey incompleteness in the construction of the random catalogues.

The probability density distributions described above are constructed in the following way. First, we optimally bin redshift distributions using the algorithm presented by Knuth (2006). This is a maximum likelihood method for determining the optimum number of bins needed to both capture the dominant features in the data and minimise the number of random sampling fluctuations. We create histograms in this way for star-forming and non star-forming galaxies separately, in magnitude bins of size 1, shifted by 0.5 magnitudes, over the range 13 to 25. We iteratively increase the magnitude bin sizes at the bright and faint ends of the magnitude distributions to ensure that there are a minimum of 20 galaxies in each redshift histogram. We then smooth the histograms with a Gaussian smoothing kernel having a standard deviation equal to the bin size, to remove spikes in the redshift distributions attributable to large-scale structure (galaxy clusters, filaments, sheets and voids). The probability density distributions are then obtained by interpolating a cubic spline over the smoothed histograms, and by normalising to the area underneath the resulting curve.

In Figure 6.1, we plot a histogram showing the redshift distribution of our total galaxy sample in blue, and the random sample divided by ϑ_g in green. The green histogram shows a smooth distribution, reflective of the overall selection function for our sample. Spikes are apparent in the real histogram attributable to large-scale structure. The smooth redshift distribution apparent in our random sample, following the overall shape of the real histogram, suggests that our approach is robust.

Random galaxies for our sample extracted from EAGLE simulation volume are distributed uniformly for each of the redshift snapshots. We created $\vartheta_g = 10$ times as many random galaxies as there are real ones.

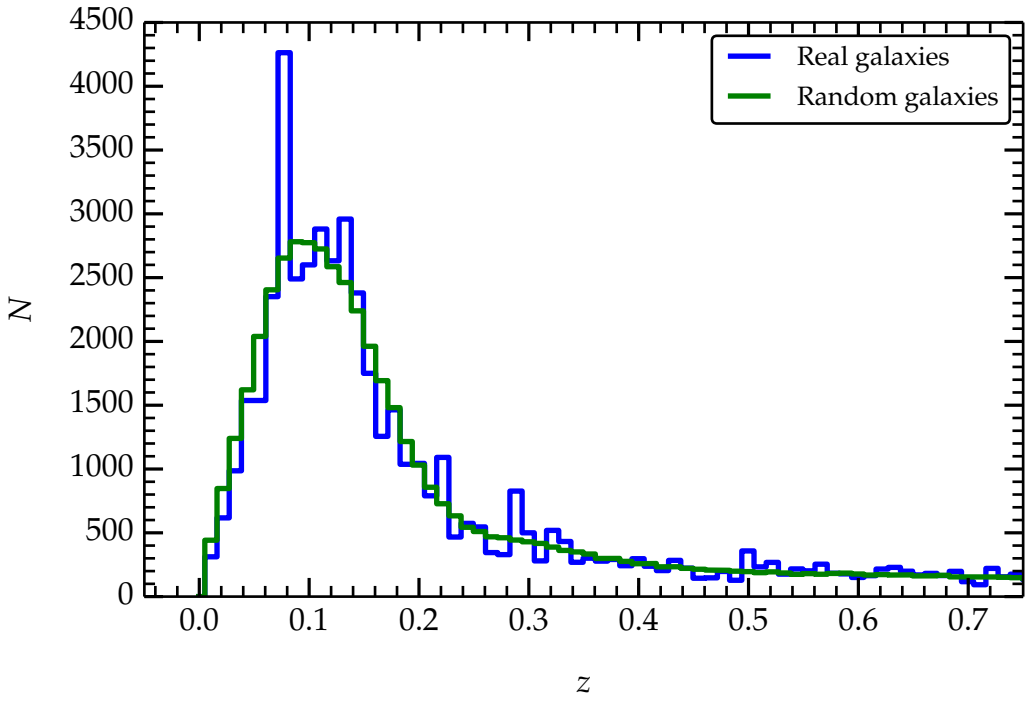

Figure 6.1: The real galaxy redshift distribution for our total sample (blue), and the random galaxy redshift distribution divided by ϑ_g (green). The smooth shape of the random distribution, following the broad shape of the real distribution, suggests that our approach to generating random catalogues of galaxies is robust.

6.4.2 Random absorber catalogues

For our observed O VI sample, which derives entirely from COS FUV spectra, we created $\vartheta_a = 1000$ random absorbers for every real one, varying the redshift, but preserving all other parameters. In this way, we randomise absorbers only along the QSO sight-lines, not transverse to them, so as to preserve the geometry of our survey. Random redshifts were chosen on the basis of an equivalent width threshold. For every real O VI absorber with Doppler broadening parameter b, and equivalent width W, we calculated the minimum equivalent width, $W_{3\sigma}$, at which the weaker transition in the doublet (O VI λ1037) for that absorber could still be observed above the required 3σ significance threshold as a function of wavelength in the spectrum from which that absorber was obtained. The significance of absorption features in COS spectra cannot be estimated in the usual way due to the non-Poissonian noise properties of the instrument (Keeney et al.,

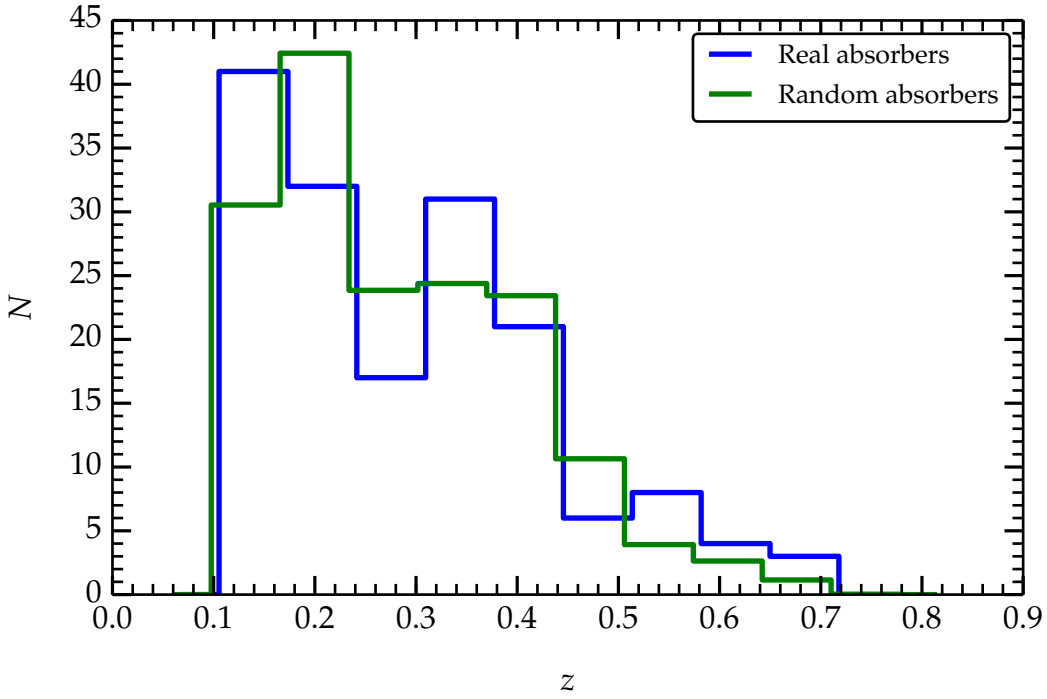

Figure 6.2: The real O VI redshift distribution for our total sample (blue), and the random O VI redshift distribution divided by ϑ_a (green). The random redshift distribution follows the same selection bias as the data, which suggests our procedure for generating random absorber catalogues is robust.

2012, see also Chapter 2, Section 2.2). Nevertheless, Keeney et al. (2012) provide a formalism for doing this, which we adopt (see their equations (4)–(5), (7) and (9)–(10)). We then transformed wavelength coordinates to redshift coordinates to obtain $W_{3\sigma}$ as a function of z, and distributed random absorbers in z where the condition $W > W_{3\sigma}(z)$ was satisfied. We enforced a maximum redshift equivalent to a -5000 km s^{-1} offset from the QSO redshift, as was done in the data.

In Figure 6.2, we show the redshift distribution of O VI absorbers for our full sample in blue in comparison to the random sample divided by ϑ_a in green. A larger number of O VI absorbers are found at lower redshifts, where they are detected in the wavelength range covered by the COS G130M grating (see Chapter 2, Section 2.2 for details). COS has greater sensitivity over this wavelength range, and the spectra from this grating therefore typically have a higher signal-to-noise ratio (SNR). The shape of the random distribution follows this closely,

which indicates our procedure is robust.

For our simulated set of O VI absorbers extracted from the EAGLE simulation, we randomise absorber velocities uniformly along the sight-lines from which they are extracted. We created $\vartheta_a = 10$ times as many random absorbers as there are real ones.

6.5 Results

In this section, we present the results of our two-point correlation analysis, following the mathematical formalism outlined in Section 6.3. The results that follow were computed using the random samples described in the previous section. We present results for our full sample of O VI absorbers and galaxies, and also for the subsamples containing only star-forming and non star-forming galaxies.

6.5.1 2D two-point correlation functions

In Figure 6.3 we show the two-dimensional correlation functions (top panels) and their uncertainties (bottom panels) for our full sample of O VI absorbers and galaxies. The results are shown in bins of 1 Mpc (comoving) and are derived from pair counts smoothed with a Gaussian kernel having a standard deviation of 1 Mpc. We justify the use of a smoothing kernel by assuming that the underlying matter distribution that gives rise to O VI absorbers and galaxies is also smooth. This approach is desirable, as it strikes a compromise between lowering the shot noise in the measurement, whilst keeping a relatively small bin size. From left to right, the panels show the O VI-galaxy cross-correlation function (ξ_{ag}), the galaxy auto-correlation function (ξ_{gg}), the O VI auto-correlation function (ξ_{aa}), and the ratio $\xi_{ag}^2/\xi_{gg}\xi_{aa}$.

On inspection of the figure, we see that on small scales, the amplitudes of ξ_{ag}, ξ_{gg}, and ξ_{aa} are highly comparable within the uncertainties. Note that we are unable to probe ξ_{aa} on transverse scales > 2 Mpc with our data. This leads to the ratio $\xi_{ag}^2/\xi_{gg}\xi_{aa}$ being close to 1 on scales $\lesssim 1$ Mpc, which suggests that O VI absorbers and galaxies are in close correspondence with one another on these

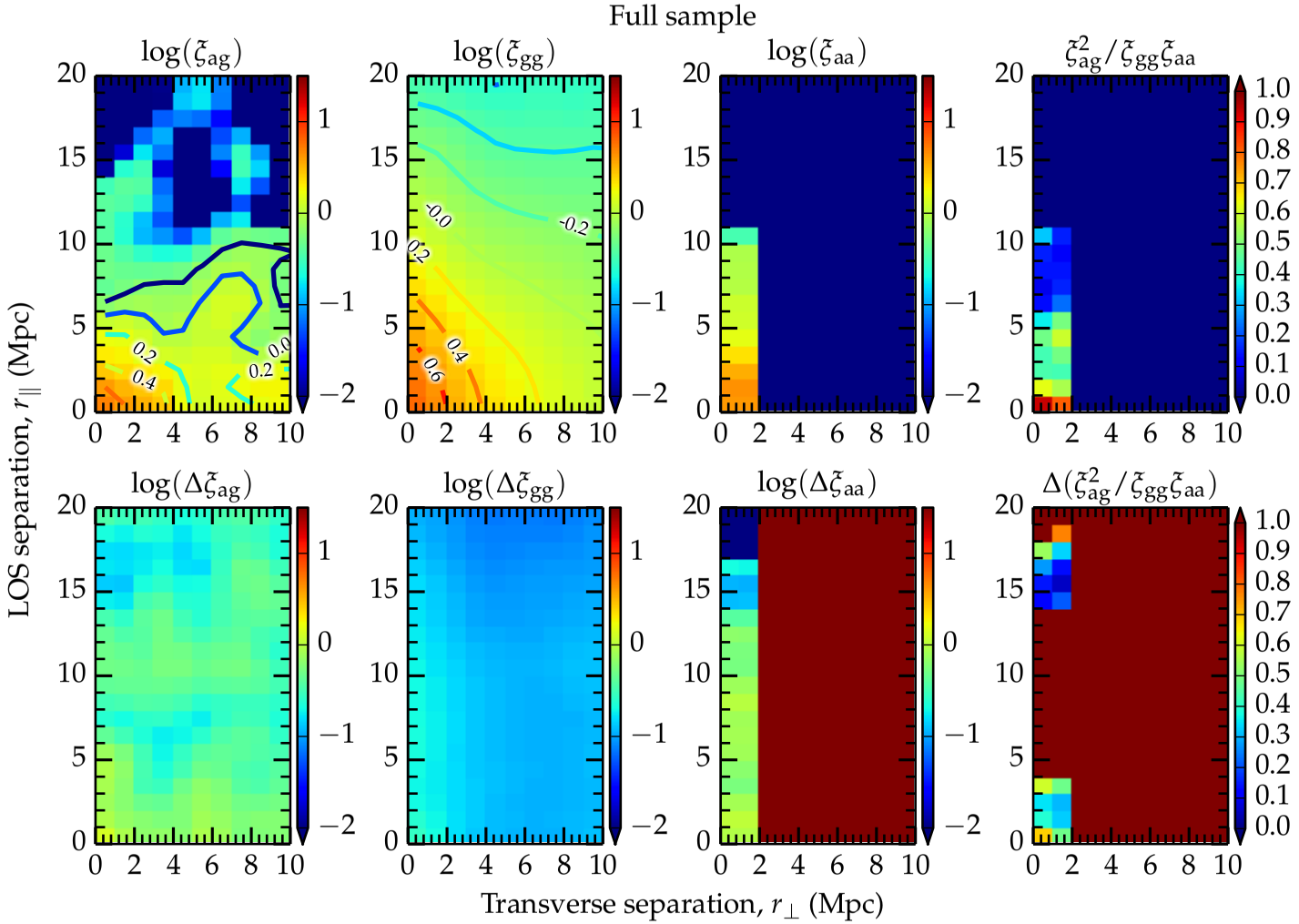

Figure 6.3: Two-dimensional correlation functions for galaxies and O VI absorption systems (top panels) and their respective uncertainties (bottom panels), as a function of separation parallel ($r_{\parallel}$) and perpendicular ($r_{\perp}$) to the LOS. From left to right: the galaxy-O VI cross correlation function (ξ_{ag}), the galaxy auto-correlation function (ξ_{gg}), the O VI auto-correlation function (ξ_{aa}), and the ratio $\xi^2_{ag}/\xi_{gg}\xi_{aa}$. Note that our data are not suitable for measuring ξ_{aa} and $\xi^2_{ag}/\xi_{gg}\xi_{aa}$ on transverse scales > 2 Mpc. The correlation functions are calculated using a bin size of 1 Mpc, and the pair-counts are smoothed with a Gaussian kernel of standard deviation 1 Mpc in both directions.

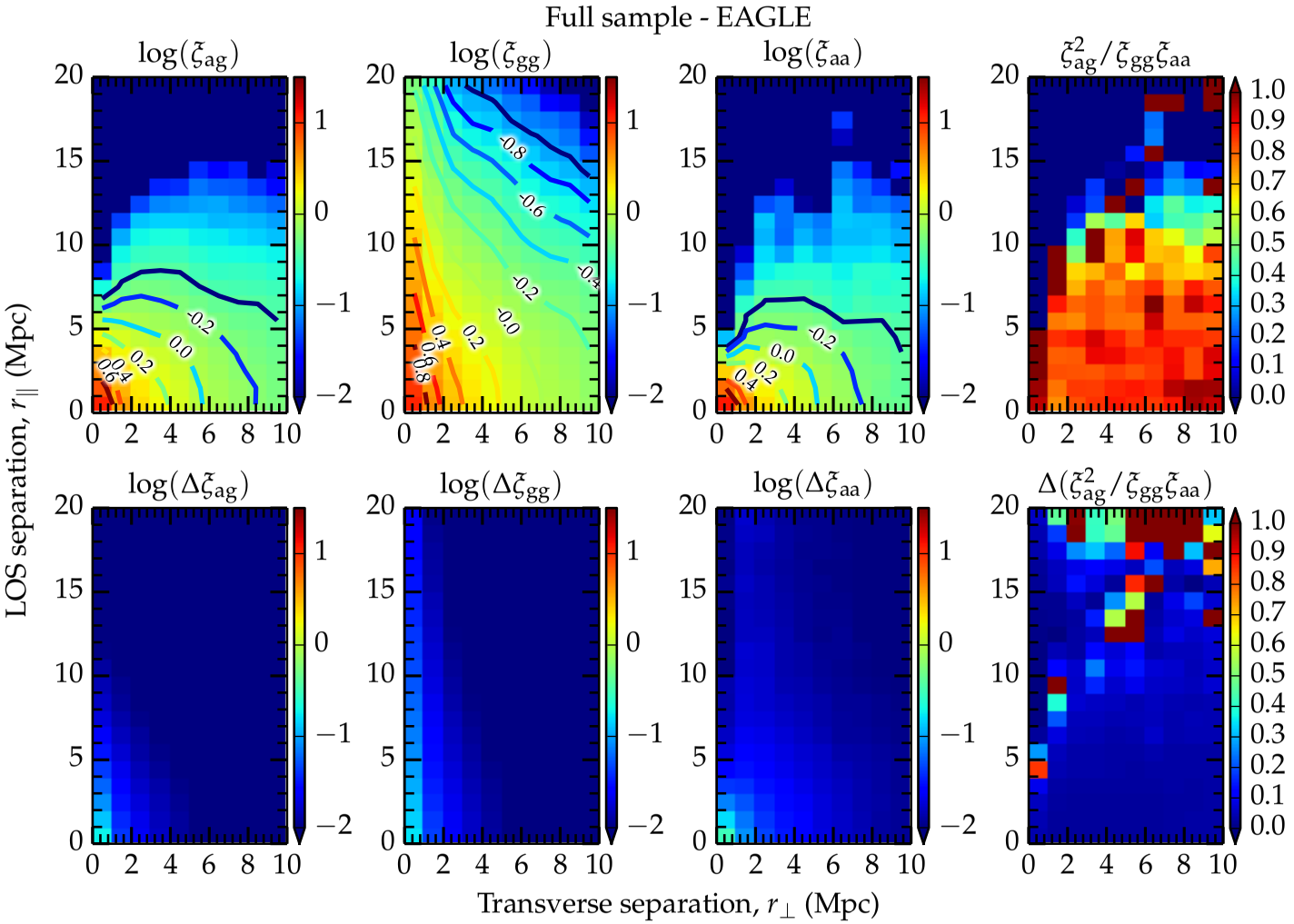

Figure 6.4: Same as Figure 6.3, but for the simulated samples extracted from the EAGLE simulation. Note that for the simulated samples we do not apply a Gaussian smoothing kernel.

scales. Our data do not have sufficient statistical power to quantify the presence of anisotropies in $\xi_{\rm ag}$ or $\xi_{\rm aa}$, but for $\xi_{\rm ag}$ we can nevertheless examine the 'isocorrelation' contours shown in these plots. We see a reasonably isotropic signal in $\xi_{\rm ag}$, at least on the small scales where our measurement is significant, which indicates that O VI absorbers show very little velocity dispersion with respect to the galaxies. There is a hint of some compression in the signal along the LOS on large scales, indicative of large-scale bulk motions (e.g. Kaiser, 1987), but not at a statistically significant level. Deviations from an isotropic signal are present in $\xi_{\rm gg}$, which we expect for a galaxy sample of the size presented here, where a significant fraction of the galaxies reside in groups and clusters, with velocity dispersions of several 100 km s^{-1}. We note that the deviation from isotropy in $\xi_{\rm gg}$ complicates the comparison of clustering amplitudes between $\xi_{\rm ag}$ and $\xi_{\rm gg}$ (and our inferences based on $\xi^2_{\rm ag}/\xi_{\rm gg}\xi_{\rm aa}$ above), since the effect of the redshift-space distortions is to 'smear' the total correlation amplitude at a given $r_\perp$ over a range in $r_\parallel$. The total correlation amplitudes at a given $r_\perp$ are therefore not necessarily as comparable as they appear in Figure 6.3. We investigate this further in the next section.

In Figure 6.4 we present the same calculation illustrated in Figure 6.3, but this time on the samples extracted from the Eagle simulation volume. For this, and all the comparisons that follow, we present results from the $z = 0.271$ Eagle snapshot, which is roughly the median redshift of our O VI sample. Note that for the Eagle calculations we don't apply a Gaussian smoothing kernel. Much like in the real case, the amplitudes of $\xi_{\rm ag}$, $\xi_{\rm gg}$, and $\xi_{\rm aa}$ are all very similar at small scales within the uncertainties. We note that the correlation amplitudes in Eagle are somewhat higher than in the data at these scales, although the Gaussian smoothing kernel employed in the latter does act to lower the correlation amplitude at small separations where they are intrinsically peaked. Inspection of an unsmoothed version of Figure 6.3 reveals that the correlation amplitudes on the smallest scales are in fact comparable to those in Eagle within the uncertainties. Also much like in the data, there is very little anisotropy on small scales in $\xi_{\rm ag}$. Even without model-fitting, it is clear that the 'anisotropy ratio' along the LOS on < 4 Mpc scales is no more than 2:1, which limits the velocity dispersion of O VI around galaxies

to $\lesssim$ 100 km s^{-1}. A highly isotropic signal is seen in ξ_{aa} on $<$ 4 Mpc scales as well, which suggests that the O VI absorbers are virtually static with respect to one another on these scales. Intruigingly, the hint of a compression along the LOS in ξ_{ag}, seen in the real data on large scales, appears in the simulated sample with high significance. The same is seen in ξ_{aa}. This then points to towards a picture in which O VI absorbers show bulk motions towards both galaxies and themselves on $\sim$ 10 Mpc scales. Given the extremely low significance of this result in the real data, we caution that this finding is far from conclusive. We note that there are some differences in ξ_{gg} between the data and the simulation. In particular, there is a larger anisotropy in the signal along the LOS in EAGLE compared to the data. However, the galaxy auto-correlation functions are not the primary focus of this study, and we shall therefore leave a more detailed comparison to future work.

Next we examine the 2D two-point correlation functions for O VI absorbers and star-forming galaxies only. This calculation for the real data is shown in Figure 6.5. Again, we see that ξ_{ag}, ξ_{gg}, and ξ_{aa} are all highly comparable within the uncertainties. This leads to a ratio $\xi_{ag}^2/\xi_{gg}\xi_{aa}$ that is highly consistent with 1 on small scales. We see no significant anisotropies in ξ_{ag} either, which suggests that on small scales, O VI absorbers and star-forming galaxies trace the same underlying distribution of matter, and show very little velocity dispersion with respect to one another (certainly $\lesssim$ 100 km s^{-1}). Redshift-space distortions in ξ_{gg} are far less evident for star-forming galaxies, and are qualitatively similar to those in ξ_{ag} and ξ_{aa}, which makes our findings based on $\xi_{ag}^2/\xi_{gg}\xi_{aa}$ somewhat more robust than those from the full sample.

In Figure 6.6 we show the same calculation in the EAGLE simulation, again with no Gaussian smoothing kernel applied. In agreement with the real data, ξ_{ag}, ξ_{gg}, and ξ_{aa} are all consistent within the uncertainties, and the ratio $\xi_{ag}^2/\xi_{gg}\xi_{aa}$ is highly consistent with 1. We see the same compression in ξ_{ag} and ξ_{aa} along the line of sight on large scales as was seen for the full sample. However, this is not detected in the real data with any significance. A small anisotropy is seen in ξ_{ag}, which amounts to a velocity dispersion between O VI absorbers and star-forming galaxies of no more than $\sim$ 100 km s^{-1}.

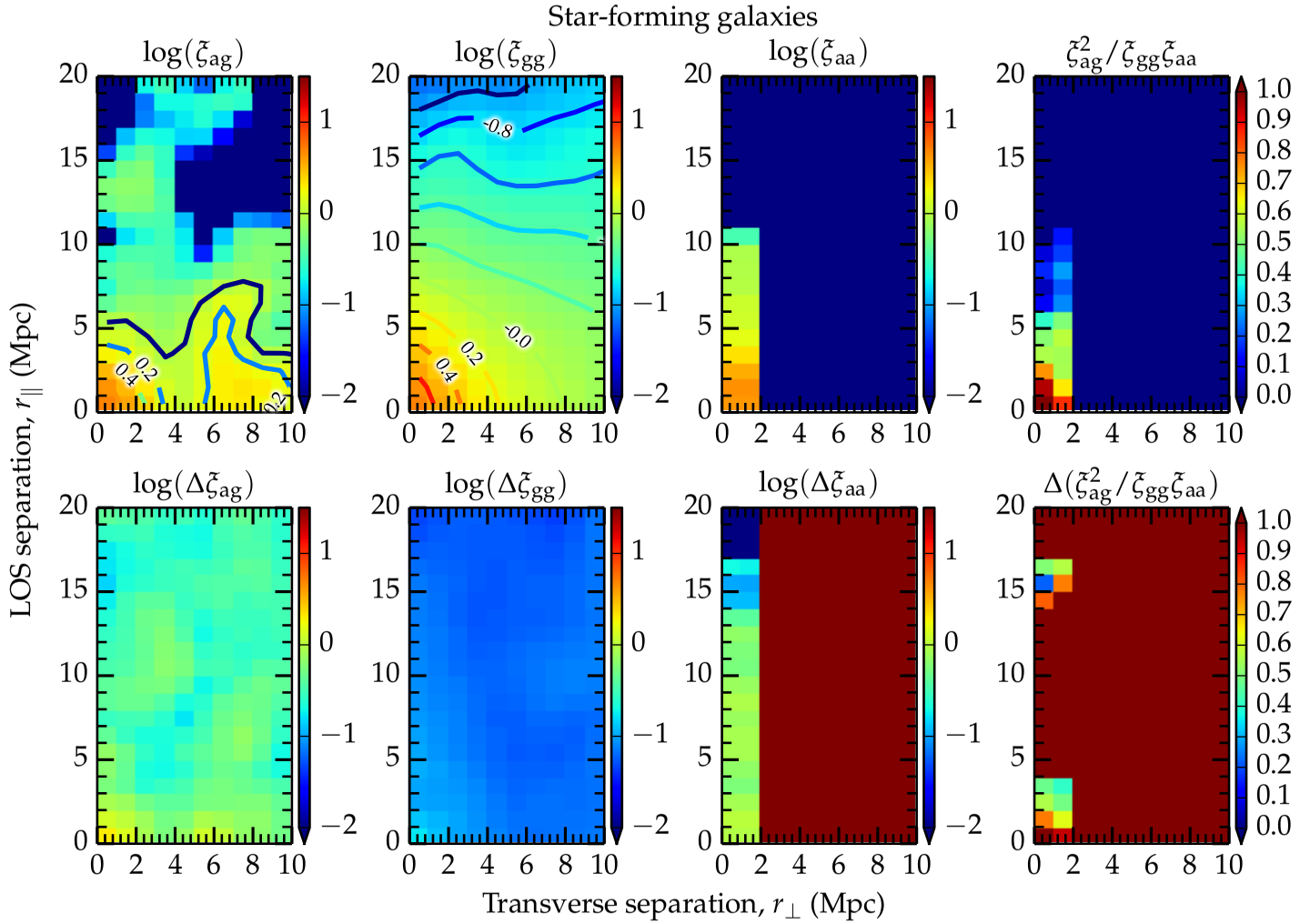

Figure 6.5: Same as Figure 6.3, but for star-forming galaxies only.

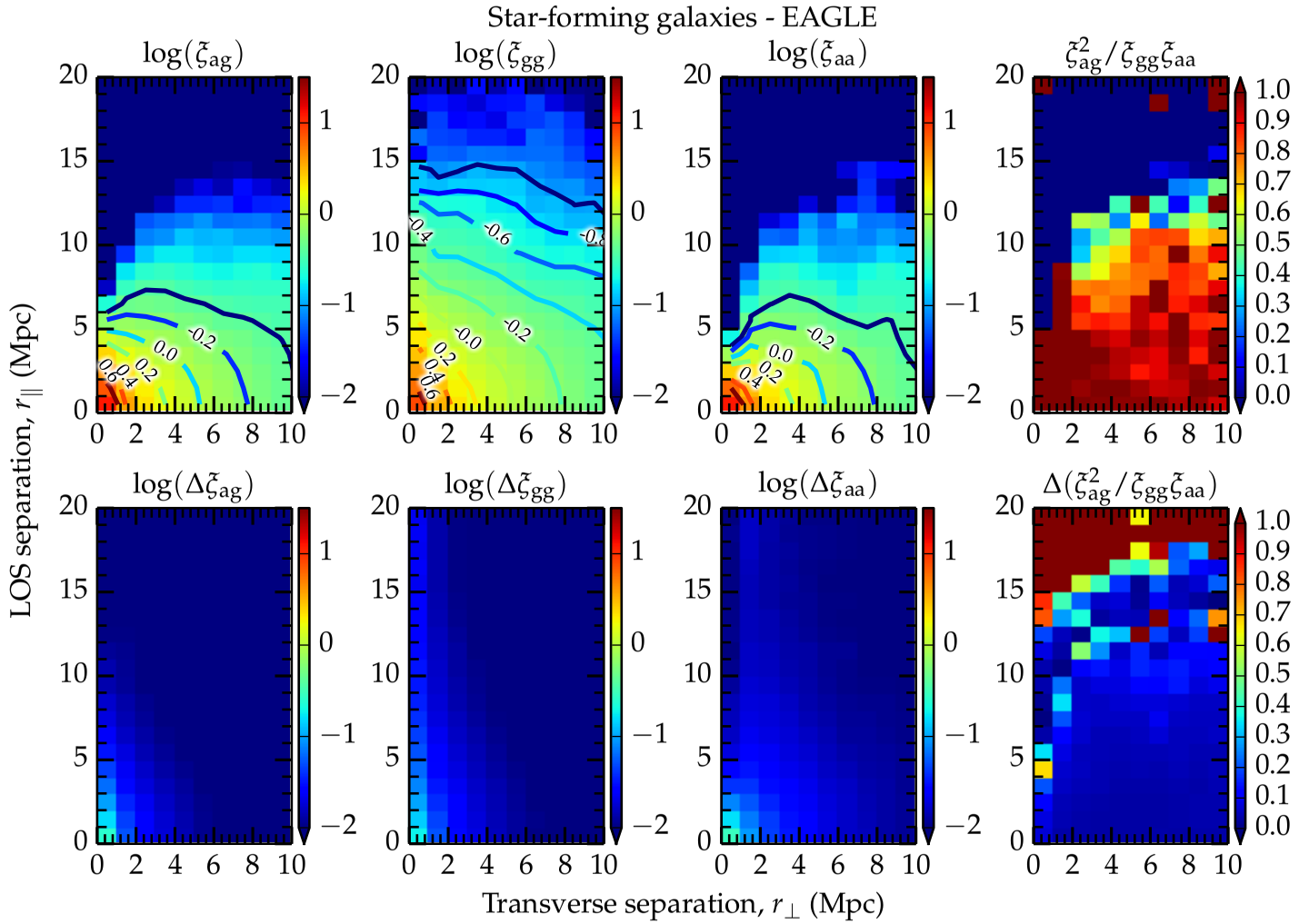

Figure 6.6: Same as Figure 6.5, but for the simulated samples extracted from the EAGLE simulation.

Finally, we examine the 2D two-point correlation functions for O VI absorbers and non star-forming galaxies only. We show this calculation for the real data in Figure 6.7. Now we see that the correlation amplitudes of ξ_{ag} and ξ_{aa} are both comparable within the uncertainties, but the amplitude of ξ_{gg} is significantly higher than both of these on small scales. As a result, we see the ratio $\xi_{ag}^2/\xi_{gg}\xi_{aa}$ is nearly consistent with zero. This suggests that O VI absorbers and non star-forming galaxies trace the underlying distribution of matter differently, although in this case we note that the interpretation of $\xi_{ag}^2/\xi_{gg}\xi_{aa}$ is complicated by redshift-space distortions that are highly-prominent in ξ_{gg}. Again, we see no evidence for any anisotropy in ξ_{ag} along the LOS, which implies that O VI absorbers show very little velocity dispersion with respect to non star-forming galaxies.

The corresponding calculation for the EAGLE simulation is shown in Figure 6.8. We see a very similar situation to the data, whereby ξ_{ag} and ξ_{aa} are both similar within the uncertainties, but the amplitude of ξ_{gg} is significantly higher than both of these. This leads to a ratio $\xi_{ag}^2/\xi_{gg}\xi_{aa}$ that is nearly zero on the smallest scales. However, we note that in both the simulation and the data, the clustering amplitudes of O VI absorbers around non star-forming galaxies are highly comparable to those around star-forming galaxies. This suggests that we should be equally as likely to find a O VI absorber close to a non-star forming galaxy as we are to find one close to a star-forming galaxy. As is consistently seen in the simulation, there does exist a small anisotropy in ξ_{ag} along the LOS, but only at the level whereby the velocity dispersion between O VI absorbers and non star-forming galaxies is $\lesssim 100$ km s^{-1}, which is broadly consistent with the real data within the uncertainties.

6.5.2 Correlation functions projected along the line-of-sight

Comparisons on the clustering amplitudes between ξ_{ag}, ξ_{gg}, and ξ_{aa} for the 2D two-point correlation functions in the previous section are complicated by the redshift-space distortions that lead to anisotropies in the signal. To better-compare the clustering amplitudes as a function of scale, we now examine the correlation functions that are projected along the LOS, as in equation (6.3.9).

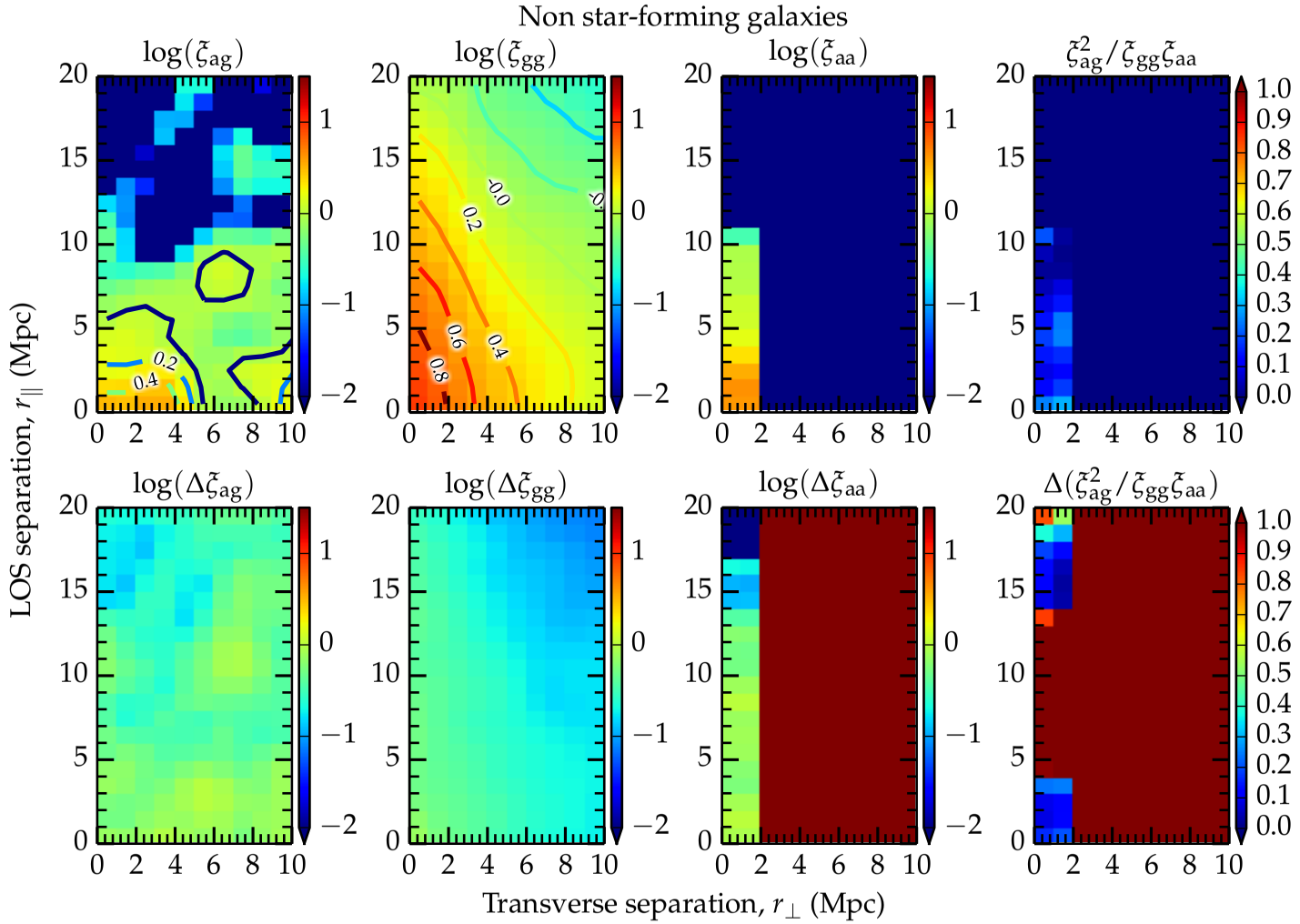

Figure 6.7: Same as Figure 6.3, but for non star-forming galaxies only.

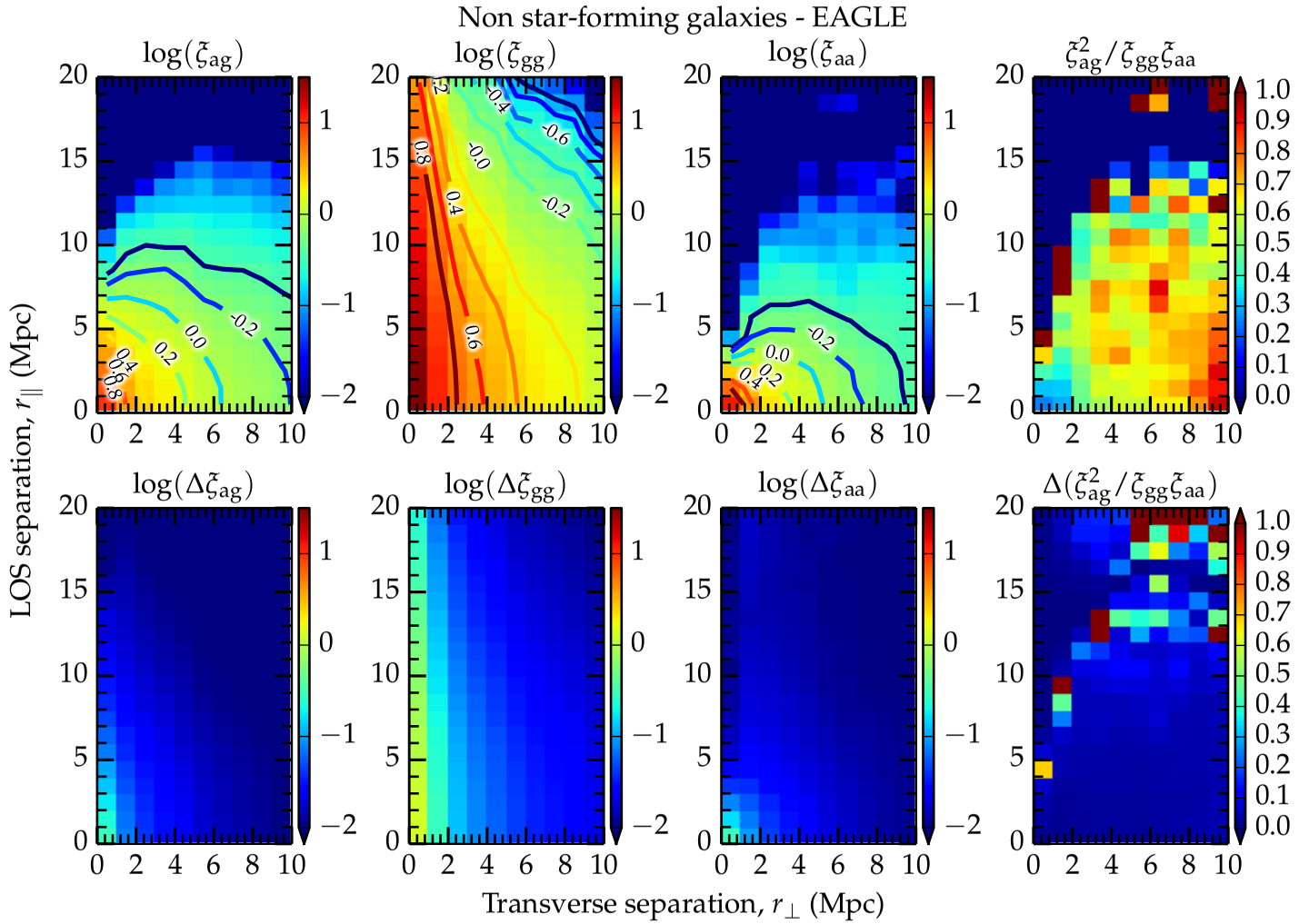

Figure 6.8: Same as Figure 6.7, but for the simulated samples extracted from the EAGLE simulation.

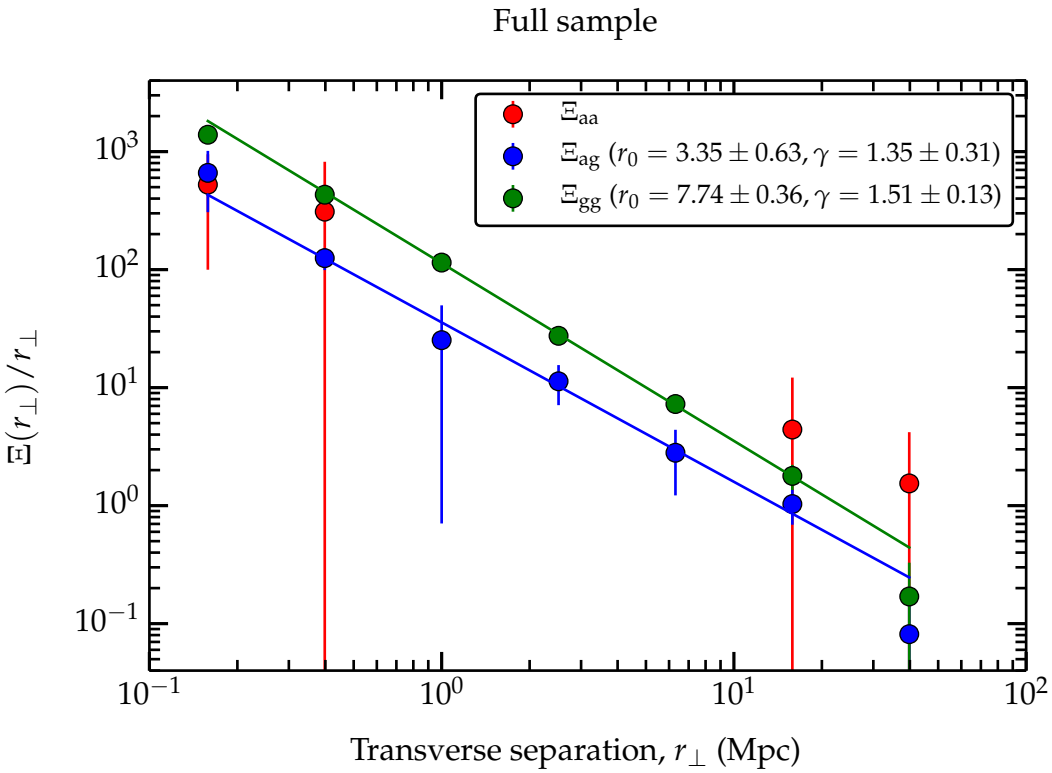

Figure 6.9: Correlation functions projected along the line-of-sight and divided by $r_\perp$, $\Xi(r_\perp)/r_\perp$, for our full sample of O VI absorbers and galaxies. Our measurement for the projected O VI-galaxy cross-correlation function, Ξ_{ag}, is shown by the blue data points. Green data points show the measurement for the projected galaxy auto-correlation function, Ξ_{gg}, and red data points show the measurement for the projected O VI auto-correlation function, Ξ_{aa}. The lines correspond to the best power-law fits (equation (6.3.12)) to Ξ_{ag}, Ξ_{gg}, and Ξ_{aa} in blue, green, and red respectively. Note that the data points and their uncertainties are correlated. Uncertainties smaller than the symbols are not shown.

In Figure 6.9, we show the correlation functions of O VI absorbers and galaxies projected along the LOS and divided by $r_\perp$, $\Xi(r_\perp)/r_\perp$, for our full sample. We show the projected O VI-galaxy cross-correlation function, Ξ_{ag}, in blue data points, the projected galaxy auto-correlation function, Ξ_{gg}, in green data points, and the projected O VI auto-correlation function, Ξ_{aa}, in red data points. For Ξ_{gg} we integrate to $r_\parallel = 45$ Mpc, and for Ξ_{ag} and Ξ_{aa} we integrate to $r_\parallel = 13$ Mpc. These integration limits are the minimum for which the data points had converged to stable values, indicating that we are fully integrating the reliably measured signal, and minimising the addition of shot noise. We note that the data points are correlated, and that uncertainties smaller than the data points are not shown. The uncertainties are bootstrap uncertainties that include both the variance and covariance in the

measurement. The lines show the best-fitting power-law slopes to the data (equation (6.3.12)) using the same colour scheme. We only fit these power-law slopes to the data at $r_\perp < 10$ Mpc, as deviations from a power-law are typically seen at $r_\perp > 10$ Mpc. We do not attempt to fit a power-law slope to Ξ_{aa}, for which the data show very low statistical significance, and are likely not robust. It is important to note that the measurement of ξ_{aa} in the 2D two-point correlation function is dominated by O VI pairs along the LOS. These pairs do not contribute to the measurement in Figure 6.9, which instead comes from transverse pairs in the very few closely-separated QSO sight-line pairs (and triples) in our sample.

Firstly, we note that a power-law is a good description of the data for Ξ_{ag} and Ξ_{gg} at $r_\perp < 10$ Mpc. We find that $\xi_{ag}(r)$ has a correlation length of $r_0^{ag} = 3.35 \pm 0.63$, and a slope of $\gamma^{ag} = 1.35 \pm 0.31$, whereas $\xi_{gg}(r)$ has a correlation length of $r_0^{gg} = 7.74 \pm 0.36$, and a slope of $\gamma^{gg} = 1.51 \pm 0.13$. We therefore find that the clustering of O VI absorbers around galaxies is less than the clustering of galaxies with themselves. The slope of Ξ_{ag} is apparently shallower than that of Ξ_{gg}, although they are consistent with one another within the uncertainties. The amplitude of Ξ_{aa}, and the apparent slope, is comparable with both Ξ_{ag} and Ξ_{gg}. The differing slopes between Ξ_{ag} and Ξ_{gg}, combined with the difference in their correlation lengths, suggests that O VI absorbers and galaxies may not linearly trace the same underlying distribution of matter in general, however this is by no means definitive given the uncertainties on the data, in particular for Ξ_{aa}.

As we did for the 2D two-point correlation functions, we now examine the correlation functions with star-forming and non star-forming galaxies separately. We show the projected correlation functions with star-forming galaxies only in Figure 6.10. Here we find that $\xi_{ag}(r)$ has a correlation length of $r_0^{ag} = 1.55 \pm 0.63$, and a slope of $\gamma^{ag} = 1.14 \pm 0.43$, whereas $\xi_{gg}(r)$ has a correlation length of $r_0^{gg} = 5.28 \pm 0.20$, and a slope of $\gamma^{gg} = 1.44 \pm 0.08$. The clustering of O VI absorbers around star-forming galaxies is therefore still less than the clustering of star-forming galaxies with themselves, and there remains marginal evidence for a difference in slope between Ξ_{ag} and Ξ_{gg}, although not at a statistically significant level. We nevertheless still find indications that O VI absorbers and star-forming galaxies may not

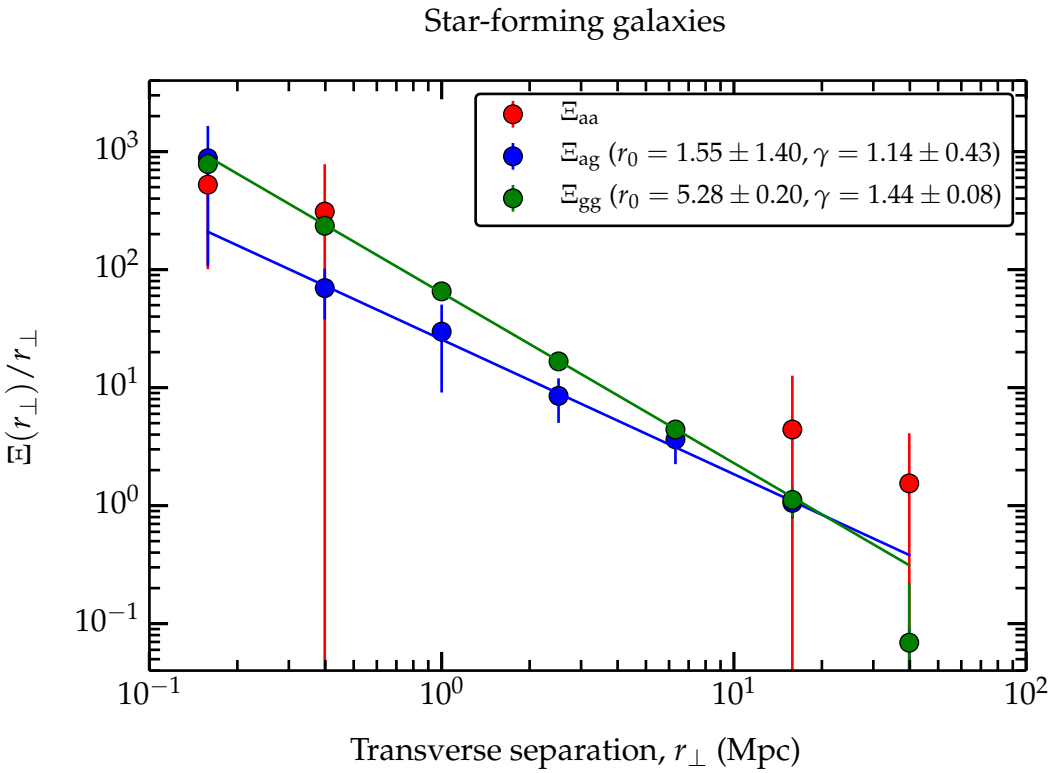

Figure 6.10: Same as Figure 6.9, but for star-forming galaxies only.

linearly trace the same underlying distribution of matter, similar to the full sample, even though the clustering amplitudes in the cross-correlation function do indicate that O VI absorbers have a strong association with star-forming galaxies in general.

Finally, in Figure 6.11, we show the projected correlation functions with non star-forming galaxies only. Here we find that $\xi_{ag}(r)$ has a correlation length of $r_0^{ag} = 2.63 \pm 1.87$, and a slope of $\gamma^{ag} = 1.25 \pm 0.71$, whereas $\xi_{gg}(r)$ has a correlation length of $r_0^{gg} = 9.61 \pm 0.47$, and a slope of $\gamma^{gg} = 1.63 \pm 0.15$. The slopes of the correlation functions are still consistent with each other within the uncertainties, but there are nevertheless indications that the slope in Ξ_{gg} is steeper than that of Ξ_{ag}. This supports indications from Figure 6.7 that O VI absorbers and non star-forming galaxies do not linearly trace the same underlying distribution of matter. It is well known that non star-forming galaxies are more biased tracers of matter than are non star-forming galaxies, as is clearly seen from the difference in the correlation lengths and slopes of their auto-correlation functions. However, the

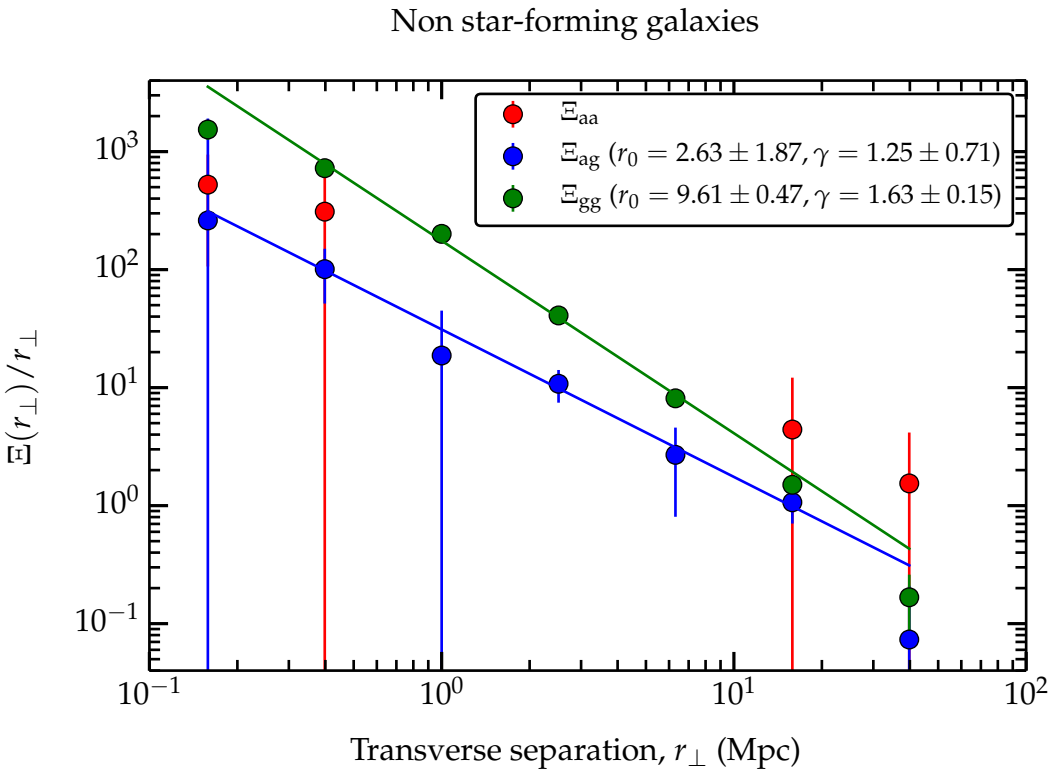

Figure 6.11: Same as Figure 6.9, but for non star-forming galaxies only.

striking result is the similarity in slopes and amplitudes of the cross-correlation functions of O VI absorbers with star-forming and non star-forming galaxies. This suggests that O VI absorbers are equally likely to be found in the same locations as star-forming galaxies as they are to be found in the same locations as non star-forming galaxies, at least over the scales probed by our data.

We now compare the correlation functions projected along the LOS to predictions from the EAGLE simulation. In Figure 6.12, we show the same measurements as in Figure 6.9 for our full sample, with the predictions from EAGLE over the redshift range $0.1 < z < 0.7$ shown as shaded regions with the same colour scheme as the data. We see that the agreement in Ξ_{ag} between the real data and that from the simulation is excellent. The simulation also produces predictions for Ξ_{aa} that are entirely consistent with the data. We find a reasonable agreement in Ξ_{gg} at $<$ Mpc scales, but this agreement diverges at larger scales due to a discernable difference in the slopes of the galaxy auto-correlation functions. We shall refrain from examining in detail the comparison between the galaxy auto-correlation functions,

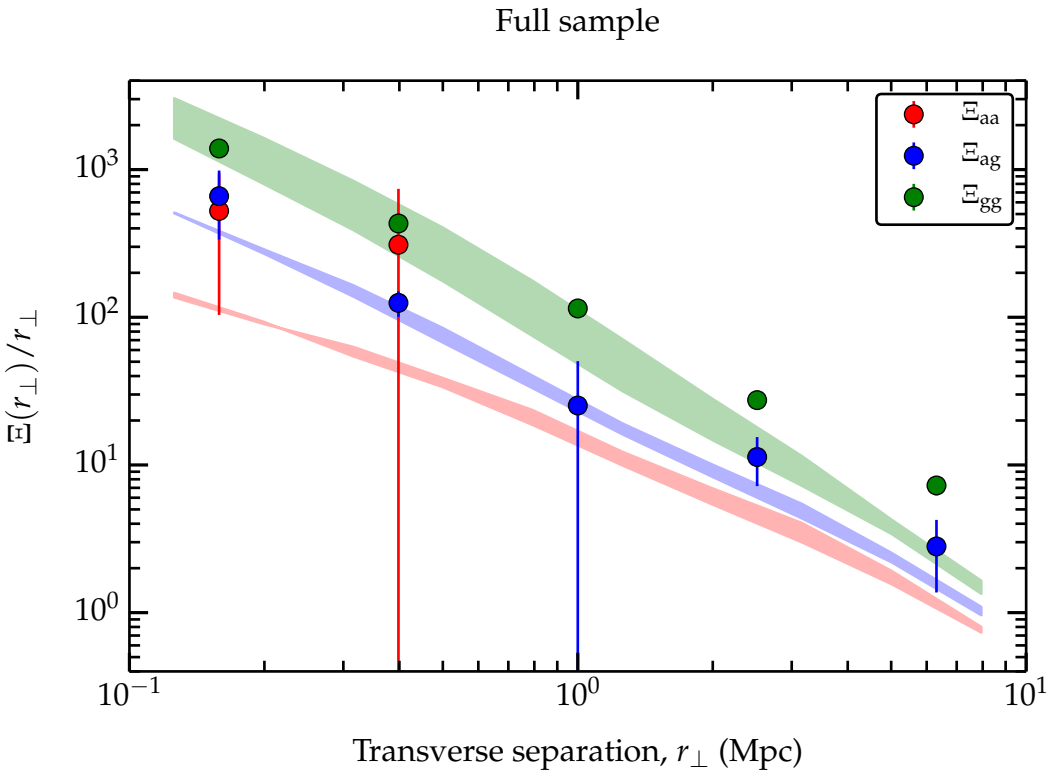

Figure 6.12: Same as Figure 6.9, but now comparing the data to the predicted slopes and amplitudes of the correlation functions from the EAGLE simulation. The shaded regions represent the predictions over the redshift range $0.1 < z < 0.7$, and adopt the same colour scheme as the data.

and leave this to future work. The relative slopes between Ξ_{gg} and Ξ_{ag} predicted by the EAGLE simulation reveal the same trend hinted at in the data, whereby the slope of Ξ_{ag} is shallower than that of Ξ_{gg}. It can also be seen that the slope of Ξ_{aa} is shallower still. This indicates that O VI absorbers and galaxies in the EAGLE simulation do not linearly trace the same underlying distribution of matter.

In Figure 6.13, we show the same comparison but for star-forming galaxies only. There is again a good agreement between the predicted and measured Ξ_{ag} within the uncertainties, albeit with a small discernable difference in the slopes of the correlation functions. An excellent agreement is seen in Figure 6.14 for the slope and amplitude of cross-correlation function of O VI absorbers with non star-forming galaxies. We also see that the amplitudes in Ξ_{ag} are in general lower than those in Ξ_{gg}, and that this is more pronounced for non star-forming galaxies, for which Ξ_{gg} is a more biased tracer of the underlying distribution of matter.

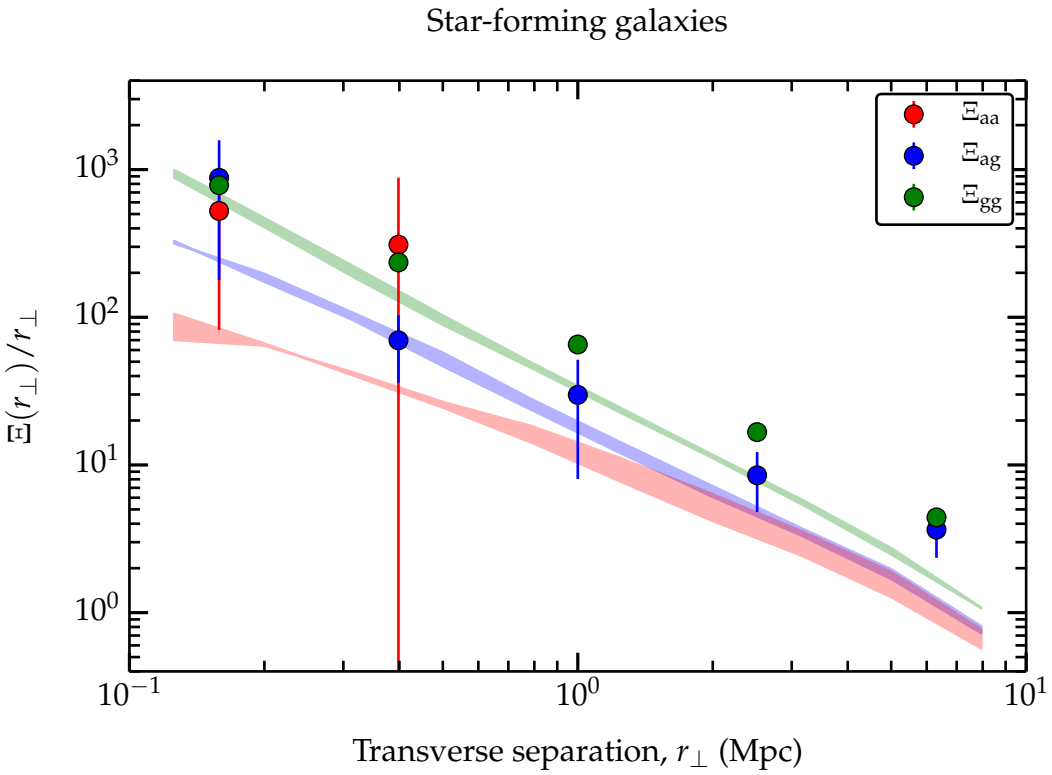

Figure 6.13: Same as Figure 6.12, but for star-forming galaxies only.

Furthermore, the predicted relative slopes in Ξ_{ag} and Ξ_{gg} show the same trends hinted at in the data, whereby Ξ_{ag} is shallower than Ξ_{gg} for both star-forming and non star-forming galaxies, with this effect being more pronounced in the case of the latter. The slope in Ξ_{aa} is also consistently shallower than both of these. We can therefore infer from the simulation that O VI absorbers and galaxies do not linearly trace the same underlying distribution of matter, and that this is most markedly the case for non star-forming galaxies. In addition, the O VI absorbers are equally likely to reside close to star-forming galaxies as they are to reside close to non star-forming galaxies. These inferences are all entirely consistent with those that can be drawn from the data.

6.6 Discussion and conclusions

In light of these new results on the cross-correlation functions of O VI absorbers and galaxies, we now explore some possible interpretations, and make compar-

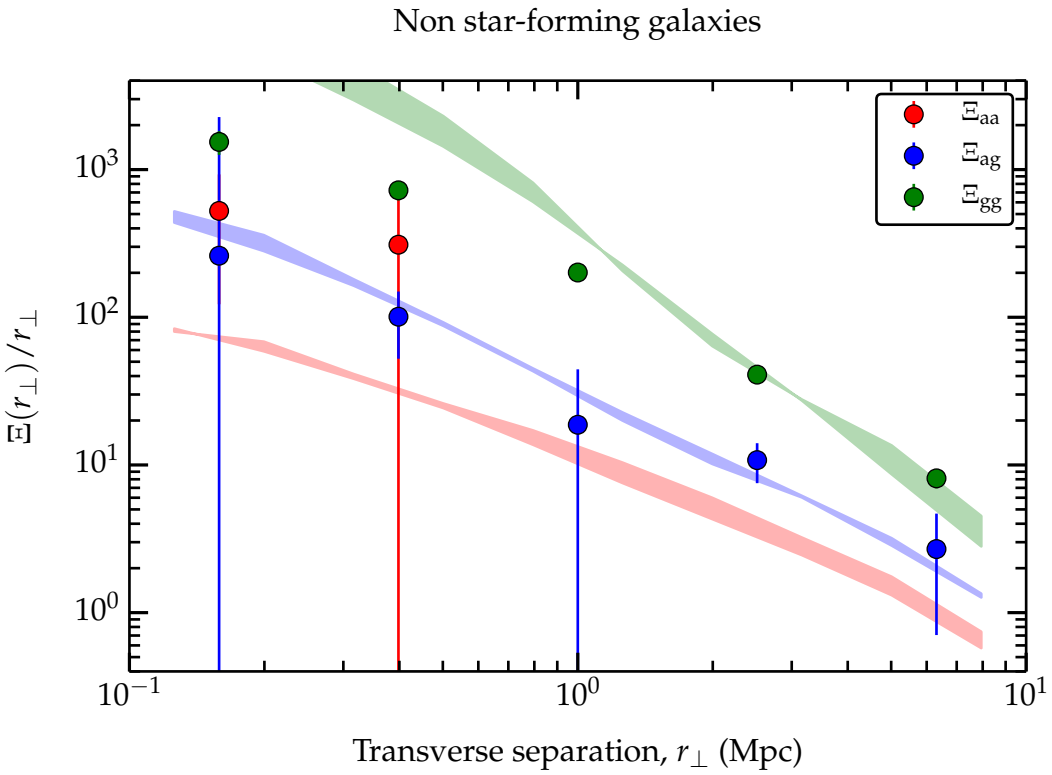

Figure 6.14: Same as Figure 6.12, but for non star-forming galaxies only.

isons to similar studies in the literature.

6.6.1 Comparisons with previous results

Chen & Mulchaey (2009) performed a very similar study to the one presented here, in which they measured the two-point cross-correlation function of O VI absorbers and galaxies projected along the LOS at $z < 1$. We note that all of the galaxy data used in that study form a small subset of the galaxy data used here, whilst their absorption-line data from the Space Telescope Imaging Spectrograph (STIS) and the Far-Ultraviolet Spectroscopic Explorer (FUSE) have been updated to that available from COS. We find that our results contrast with these earlier results in two primary aspects: (i) Chen & Mulchaey (2009) find that the clustering amplitudes of O VI absorbers around star-forming galaxies are comparable to those of star-forming galaxies with themselves, whilst we find that in general they are smaller; and (ii) Chen & Mulchaey (2009) find that the clustering amplitudes of O VI absorbers around star-forming galaxies are weaker than those

around non star-forming galaxies, whilst we find that they are comparable. To explain these differences it is important to note the sample sizes. Chen & Mulchaey (2009) used a sample of 13 O VI absorbers and 670 galaxies, which is substantially smaller than the sample assembled for this study. The quoted uncertainties on their measurements are also Poissonian, which underestimates the true uncertainties (see Section 6.3). We therefore find that our measurements are almost certainly more statistically robust, and that with more conservative estimates on the uncertainties, the results between these studies may in fact be consistent.

Turner et al. (2014) performed a study that examined the distribution of O VI, H I, and other metal ions around star-forming galaxies at $z \approx 2.4$. Their study uses a different technique to ours, in which they measure the median optical depth of O VI in spectral pixels binned in terms of their transverse and LOS separation from galaxies. Their study is optimised to investigate the association between absorbers and high-redshift galaxies on small (< Mpc) scales, whereas ours is optimised to investigate the association between absorbers and galaxies at $z < 1$ on predominantly large scales, spanning the range $0.1 \lesssim r_{\perp} \lesssim 10$ Mpc. Given the difference in approach between these studies, our comparison is restricted to being purely qualitative. Nevertheless, we note that the distribution of O VI around galaxies presented in Turner et al. (2014) reveals stronger anisotropies along the LOS than are evident in our study. It is important to note that these anisotropies are revealed on scales of a few hundred (proper) kpc, which are substantially smaller than the $\sim$ Mpc scales (comoving) that are considered here. It will be interesting to investigate the evolution in the distribution and dynamics of O VI around galaxies from high redshifts, before the peak in star-formation activity at $z \sim 2$, to the present day, but there are no straight-forward comparisons at present.

Prochaska et al. (2011b) investigated the incidence of H I and O VI absorbers around galaxies at $z \lesssim 0.5$. Note that we make use of a subset of their galaxy redshift data in this study (Prochaska et al., 2011a). By examining the incidence rate of O VI absorbers as a function of galaxy impact parameter, they infer that the covering fraction of O VI around sub-$L^{\star}$ galaxies is nearly unity to impact parameters of 300 kpc. Comparing to the total incidence rate per absorption path length, they

conclude that the majority of O VI absorbers to an equivalent width limit $W > 0.03$ Å in O VI λ1031 arise within 300 kpc of sub-$L^\star$ galaxies in the 'extended CGM', and that they rarely originate in the WHIM predicted by cosmological hydrodynamical simulations (e.g. Cen & Ostriker, 1999; Davé et al., 2001). The strategy of Prochaska et al. (2011b) is (i) to identify galaxies at small (< 300 kpc) impact parameters from the QSO sight-lines in their sample observed with the Goddard High Resolution Spectrograph (GHRS), STIS, and FUSE, then to search for O VI absorbers close to those sight-lines; and (ii) to search for galaxies close to particular O VI absorbers. In the latter, they find that 5 out of the 30 O VI absorbers do not have a galaxy within 1 Mpc, although this result is sensitive to the completeness limit of the galaxy survey, which may miss low luminosity galaxies. Our results are robust to galaxy completeness, and indicate that the clustering amplitudes of O VI absorbers with galaxies are weaker than those of galaxies with themselves. This may imply that not all O VI absorbers can be found close to galaxies, or that covering factors of O VI around galaxies on the scales probed ($\gtrsim$ 100 kpc) are less than 100% (or both). The conclusions of Prochaska et al. (2011b) are based on assumed covering factors of O VI around sub-$L^\star$ galaxies of close to 100% to 300 kpc, which highlights a potential source of tension with our results. Similar studies to Prochaska et al. (2011b), e.g. Tumlinson & Fang (2005) and Stocke et al. (2013), suggest that the covering fractions are significantly below 100%. We suggest that studies such as these may need to probe galaxies to fainter luminosities in order to reconcile differing results that may be biased by the galaxy survey completeness limits.

Tumlinson et al. (2011) performed another study on the incidence of O VI absorbers around galaxies, and found a dichotomy between the incidence rate around star-forming galaxies and non star-forming galaxies. In particular, they find that O VI absorbers are nearly ubiquitous within 150 kpc of star-forming galaxies, but only a small fraction of non-star forming galaxies show O VI absorption within this distance. This result contrasts strongly with ours, in which we find that the presence of O VI around star-forming galaxies is equally as likely as it is around non star-forming galaxies. However, it is important to note that

our study is optimised for scales $\gtrsim$ 100 kpc away from galaxies, and so if this dichotomy only exists on the smallest scales, we would not have been able to detect it.

6.6.2 Interpretation of the results

The principal findings of this study are as follows:

1. O VI absorbers show little velocity dispersion ($\lesssim$ 100 km s^{-1}) with respect to galaxies on $\sim$ Mpc scales.

2. The correlation length and (potentially) the slope of the O VI-galaxy cross-correlation function is weaker than that of the galaxy auto-correlation function in general, which indicates that O VI absorbers and galaxies may not linearly trace the same underlying distribution of matter.

3. On scales $\gtrsim$ 100 kpc, O VI absorbers are equally as likely to reside in the same regions as star-forming galaxies as they are to reside in the same regions as non star-forming galaxies.

Given that the enrichment of the IGM is likely attributable to galaxy feedback, it is convenient to think of these results from a galaxy-centric viewpoint, as follows.

In item 2 from the list above, the lower correlation amplitudes in the O VI-galaxy cross-correlation functions compared to the galaxy auto-correlation functions imply that either O VI absorbers are not ubiquitous to galaxies, or that their distribution around them is patchy, i.e. the covering factor of O VI around galaxies is substantially less than 100%, as was suggested in the previous section. In reality, both of these inferences could be true. This may be a function of O VI column density and/or Doppler broadening parameter, although we have not attempted to split the O VI absorber sample in this study due to low number statistics. We also note that this conclusion may not necessarily apply to lower column density O VI absorbers, below the detection limits of the present survey.

Also in item 2, indications of a shallower slope in the O VI-galaxy cross correlation function compared to the galaxy auto-correlation function could indicate

that the distribution of O VI around galaxies is more extended than the distribution of galaxies around themselves in general. From the data, we cannot rule out the possibility that O VI absorbers are primarily attached to galaxies, and that the cross-correlation amplitudes are primarily driven by clustering of the galaxies with themselves. Nevertheless, the inference of an extended distribution of O VI around galaxies is supported by predictions from the EAGLE simulation, which also indicate a shallower slope in the O VI-galaxy cross-correlation function, and an even shallower slope in the O VI auto-correlation function. The difference in the slopes and amplitudes of these correlation functions leads to the possibility that O VI absorbers and galaxies do not linearly trace the same underlying distribution of matter, and that some O VI absorbers may be found far from galaxies. This picture agrees well with the inferences made by Stocke et al. (2013), who find that broad O VI absorbers may trace hot gas ($T \approx 10^6$ K) that extends to large distances ($\sim 400 - 600$ kpc) around galaxies.

This same scenario is seen in cosmological hydrodynamical simulations, such as those presented in Hummels et al. (2013) and Ford et al. (2013), where O VI is distributed far from galaxies with a relatively flat radial profile. Comparisons made here with the EAGLE simulation suggest that the simulations are capable of reproducing the observed distribution of O VI absorbers around galaxies, as inferred from the slope and amplitude of the O VI-galaxy two-point cross-correlation function. These results also raise the tantalising possibility that a fraction of O VI absorbers do in fact arise in the WHIM, outside of galaxy haloes and groups at temperatures of $10^5 < T < 10^7$ K, as predicted by the simulations. Nevertheless, a targeted approach to detecting the WHIM is still needed if we are to confirm these predictions, and it is not yet clear whether the commonly assumed tracers of the WHIM (O VI, Ne VIII, broad Lyα) trace the bulk of this hot intergalactic plasma (e.g. Tepper-García et al., 2011, 2012, 2013).

In many respects, the outstanding agreement on the clustering of O VI absorbers around galaxies between the EAGLE simulation and the real data is quite surprising when we consider the potential origins the O VI absorbers, some of which are expected to arise from conductive and turbulent interfaces (see Chap-

ter 2, Section 2.8 for a discussion) that are clearly not resolved by the cosmological simulations at present. Either these processes are relatively unimportant for the overall population of O VI absorbers detected by current instrumentation, or they are co-spatial with other production mechanisms (e.g. photoionization, collisional ionization, shocks), the physics of which are well captured by the simulations. Given the level of agreement seen here, we also suggest that the shortcomings of our approach to extracting O VI absorbers from the EAGLE simulation, described in Chapter 5, Section 5.3.2, have little overall effect in the measured clustering signal of O VI around galaxies, although it will be important in future work to verify our results with a rigorous Voigt profile fitting procedure.

Item 1 in our list of findings from the present study indicates that we have not found substantial evidence for gas outflows or inflows traced by O VI around galaxies on $\sim$ Mpc scales at low redshifts. This scenario is also consistent with that predicted by the EAGLE simulation, which implements subgrid prescriptions for effective feedback from supernovae and AGN in order to match the present day statistics of the galaxy population. Our constraints on the velocity dispersion of O VI around galaxies are consistent with a scenario in which the majority of the O VI absorbers detected with current instrumentation move with the galaxies, and that those bound to galaxy haloes are not moving with velocities sufficient to escape their local gravitational potential. This is then suggestive of a scenario in which the wider IGM not bound to individual galaxy haloes may have been enriched early in the history of the Universe (e.g. Wiersma et al., 2010).

Our other main finding indicates that the presence of O VI absorbers on $\gtrsim 100$ kpc scales around galaxies is in no way biased towards whether those galaxies are star-forming or not. This situation is clearly echoed in the EAGLE simulation also, and indicates that the instantaneous star-formation activity in galaxies bears no relation to the overall distribution of metals around them. This further supports the inference that a significant proportion of the metals in the IGM have been distributed into the IGM early, and that the extent of ongoing star-formation has no discernible effect on the metal enrichment of the IGM on $\gtrsim 100$ kpc scales.

Importantly for this study, we have found that the dynamics and distribution of O VI absorbers around galaxies as inferred from the clustering of one population relative to the other is very well reproduced by the EAGLE simulation. This prediction is made blind, since the simulation is not calibrated in any way to match these observational results. We therefore suggest that the subgrid physics implemented in cosmological hydrodynamical simulations is sufficient for the purposes of making predictions on the dynamics and distribution of metal-enriched gas around galaxies, which represents a crucial element in our understanding on their formation and evolution.

Chapter 7

Summary and Conclusions

7.1 Key findings

In this thesis, we have explored the relationship between the metal-enriched intergalactic medium (IGM) and galaxies at $z < 1$. Galaxies have been surveyed in emission with optical spectroscopy, while the IGM has been surveyed by means of ultraviolet (UV) quasar (QSO) absorption-line spectroscopy. We have presented observational results on a complex absorption system associated to a QSO, and addressed some key questions regarding the nature of feedback in QSO host galaxies. We have also conducted a statistical study on the relationship between O VI absorbers and galaxies via measurements of their two-point cross- and auto-correlation functions.

The main methods for this work were outlined in three chapters:

- In Chapter 2 we outlined the methods for constructing a survey of the IGM along QSO sight-lines. UV spectroscopy is required to provide coverage of the key atomic transitions responsible for absorption lines attributable to intergalactic gas at low redshifts. We therefore made use of the Cosmic Origins Spectrograph (COS) on the Hubble Space Telescope (*HST*), which has a sensitivity in the far ultraviolet (FUV) more than ten times that of the previous generation UV spectrograph. We performed a full data reduction, pseudo-continuum normalisation, and Voigt profile decomposition on 7 QSOs, and obtained similar measurements from the literature for 53 more. This set of measurements, consisting of line identifications, column densities, Doppler broadening parameters and redshifts, formed our observational survey of

the IGM. We concluded this section by presenting some simple statistics on the O VI absorber sample, which was the focus of our analysis later.

- In Chapter 4 we presented our combined survey of galaxies around the QSO sight-lines presented in Chapter 2. Some of these were specifically targeted around those sight-lines, while others provided serendipitous coverage. Many properties of the galaxy sample were obtained from measurements in the literature, primarily their redshifts, magnitudes, and spectral types. We described procedures for creating a unified redshift confidence labelling across our sample, and for ensuring the best possible photometric and astrometric calibration. We also outlined spectral line measurements conducted on a subset of our sample, for which spectral type classifications were not available, or poorly determined previously. Overall, we were left with a set of 107 149 galaxies having secure redshifts, and many also with multi-band photometry and spectral type classifications. This then formed our base galaxy sample.

- In Chapter 5, we outlined procedures for generating a comparison set of data from the Evolution and Assembly of GaLaxies and their Environments (EAGLE) suite of cosmological hydrodynamical simulations. We used snapshots of the reference 100 Mpc volume over the redshift range $0.1 \lesssim z \lesssim 0.7$ to construct our comparison sample. O VI absorbers were extracted from the simulation along randomly placed one-dimensional sight-lines to mimic the observational procedure. We performed measurements of their column densities and Doppler broadening parameters using the optical depth distribution along those sight-lines, and the gas temperatures inferred from the simulation. A subset was then extracted with the same selection biases present in the observational sample. We also created a galaxy sample, identified on the basis of saddle points in the density distribution of particles within friends-of-friends haloes. Stellar masses were evaluated within circular apertures designed to mimic the observational procedure, and a mass cut was applied to reflect the selection biases in the data. This formed the

simulated comparison set central to much of the analysis in Chapter 6.

The main results in this thesis came from two primary studies. In Chapter 3, our analysis on the complex absorption profiles associated to the QSO FBQS J0209-0438 (Q0209) at $z \sim 1$ resulted in the following primary findings:

1. The associated absorption line (AAL) system in Q0209 has complex velocity structure, spanning an overall width of ≈ 600 km s^{-1}, and traces gas that is largely outflowing at several 100 km s^{-1}. This is relatively small compared to many AALs that show outflow velocities of several thousand km s^{-1}. It is not yet clear whether the broad absorption lines (BALs) trace a different aspect of the same phenomenon, with outflow velocities that often exceed 10^4 km s^{-1}.

2. The outflowing material is constrained to lie at a distance of $2.3 \lesssim R \lesssim 6.0$ kpc from the centre of Q0209, which appears to be typical for the highly-ionized AALs detected in COS FUV spectra to date.

3. The absorber covering fractions are less than unity across all ions, and show little variation. This implies that the absorbing structures have transverse sizes $\lesssim 10^{-2.5}$ pc.

4. Photoionization equilibrium models suggest ionization parameters covering around two orders of magnitude, and absorption path lengths covering around four orders of magnitude. This, combined with the inferred transverse cloud sizes, poses a challenge to models involving simple cloud geometries for associated systems. Our data are also difficult to reconcile with one, or even two-phase ionization models.

5. Collisional ionization equilibrium is effectively ruled out as a possible ionization mechanism for producing the AALs.

6. The inferred (small) cloud sizes, together with the unprecedented range in ionization parameter, suggest the absorbers may be far from both dynamical and ionization equilibrium. Simple numerical experiments indicate that

repeat observations may be able to pin down a non-equilibrium ionization solution to the data, which would naturally be explained within a framework involving continuous in-situ cloud formation in passing shock waves, and dynamical expansion of the gas.

7. The gas outflow velocities and densities inferred from the AALs, together with the distance measurement, are all consistent with scenarios involving gas entrainment or condensations in winds driven by either supernovae, or the super massive black hole (SMBH) accretion disc. If the AALs arise from the latter, then kinematic arguments suggest they may be entrained in a hot wind not detected in the UV, that would trace the majority of the mass in the outflow. We speculate that this wind may be detectable as continuous bound-free absorption in X-rays.

In Chapter 6, our analysis on the two-point cross- and auto-correlation functions of O VI absorbers and galaxies at $z < 1$ resulted in the findings outlined below:

1. O VI absorbers show very little velocity dispersion with respect to galaxies on $\sim$ Mpc scales at low redshifts. We estimate that this velocity dispersion amounts to $\lesssim 100$ km s^{-1}.

2. The slope of the O VI-galaxy cross-correlation function is potentially shallower than that of the galaxy auto-correlation function in general. We therefore find that these populations may not linearly trace the same underlying distribution of matter. In particular, these results indicate that the distribution of O VI around galaxies could be more extended than the distribution of galaxies around themselves. We therefore speculate that a fraction of the O VI absorbers might trace the warm-hot intergalactic medium (WHIM) predicted by cosmological hydrodynamical simulations.

3. The clustering amplitudes of O VI absorbers around star-forming galaxies are highly comparable to those around non star-forming galaxies. We therefore find that O VI absorbers are equally likely to inhabit the same regions typically occupied by star forming galaxies as they are to inhabit the same

regions typically occupied by non star-forming galaxies, at least on scales $\gtrsim 100$ kpc.

4. The amplitude of the O VI-galaxy cross-correlation is typically lower than that of the galaxy auto-correlation function by factors of a few. This indicates that O VI absorbers are either not ubiquitous to galaxies, or that their distribution around them is patchy on scales $\gtrsim 100$ kpc (or both), at least for the column densities at which most are currently detected.

5. We find that predictions from the EAGLE cosmological hydrodynamical simulation are remarkably consistent with the observational findings outlined above. This suggests that simulations such as these may be regarded as a powerful tool for understanding the distribution and dynamics of metal-enriched gas in the Universe.

7.2 Future prospects and concluding remarks

In Chapters 3 and 6, we have presented specific prospects for future work relating to the study of AALs in QSO spectra, and the clustering of gas around galaxies respectively. Here we conclude this thesis by taking a step back to look at the broader perspective.

In our theoretical understanding of galaxies and large-scale structure in the Universe, a clear picture is emerging. In this picture, the gas flows in and around galaxies are some of the most important processes governing their formation and evolution. Feedback from supernovae and active galactic nuclei (AGN) regulate the process of star formation, which is fuelled by filamentary gas flows and gas condensations from hot coronal reservoirs. These processes are thought to drive the galaxy population towards the global properties we see today, in statistics such as the galaxy stellar mass function. To advance our understanding on the formation of galaxies and large-scale structures, we therefore need to obtain detailed observational constraints on the interplay between galaxies, and the gaseous environments in which they reside.

Observations of the IGM and circumgalactic medium (CGM) are now providing crucial information on, for example, the distribution, dynamics, metal enrichment, temperature, and ionization state in the gas surrounding galaxies. The work presented in this thesis is just one contribution to our pool of knowledge in this area. Nevertheless, key questions remain. In particular, 30 – 40% of the baryons in the low-redshift IGM are still not well accounted for observationally, with the primary indication from theory being that they should reside in the so-called WHIM, at temperatures in the range $10^5 < T < 10^7$ K. There are some claims in the literature for a detection of the WHIM, but a complete census remains out of reach. It is also not yet clear to what extent the commonly used tracers (e.g. O VI, Ne VIII) account for this gas. The existence of the WHIM is a clear prediction from numerical galaxy formation models that implement effective feedback to reproduce the present day statistics of the galaxy population. Identifying the WHIM is therefore a clear priority if we are to critically test these models further. Future experiments may need to adopt a targeted approach to identifying the WHIM, for example by investigating the gaseous environments of spiral-rich galaxy groups, where the WHIM is expected to dominate the baryon budget.

The dynamics of gas around galaxies remains poorly constrained observationally. In particular, separating infalling gas from outflowing gas in the regions around galaxies remains a key observational challenge. Detecting gas outflows 'down-the-barrel' in the absorption and emission spectra of AGN and galaxies has proved to be a fruitful approach, and studies like the one in this thesis are revealing a highly complex and multi-faceted picture. Nevertheless, it remains hard to assess the effectiveness of this feedback in regulating the process of star formation, and enriching the IGM with metals. An unanswered question is *when* were the metals released into the IGM? Did it predominantly happen early in the Universe's history, or is it still ongoing? What fraction of the outflows follow ballistic trajectories, and what fraction escape? Predictions are now available from numerical simulations, but it still remains to test whether these predictions are reasonable. We have made steps towards that goal in this thesis, but much work is still required in this area.

Detecting the filamentary accretion flows ubiquitous in hydrodynamical simulations is more difficult than detecting outflows due to the complex geometry of the problem, and a lack of information tying the IGM dynamics to the galaxy gas dynamics. A future avenue of approach will be to combine deep, spatially resolved spectroscopy of galaxies and their immediate environments with QSO absorption-line spectroscopy, which should help tie together the dynamics on small scales with the dynamics on large scales. Another approach would be to attempt IGM tomography via 3D reconstruction from multiple QSO and/or galaxy sight-lines. The space-density of QSOs alone is in general too low to make this goal achievable, and the use of distant galaxies for this kind of work is highly challenging with current instrumentation. Nevertheless, it promises to be a powerful technique in the future.

Finally, the sources of feedback required by numerical simulations of galaxy formation remain poorly understood. Various feedback mechanisms from both AGN and supernovae have been proposed in the literature, but many lack fundamental observational constraints. The relative importance of these feedback processes in the context of galaxy evolution is also not well determined. Better observational constraints and numerical modelling techniques will undoubtedly be required in the future, if we are to better understand the key aspects of feedback.

Many new questions have arisen from the work presented in this thesis, but our work undoubtedly provides a modest and yet valuable contribution to our understanding of the relationship between the IGM and galaxies, and the evolution of baryonic structure in the Universe.

Charles Finn

Durham 2015

Appendix A

A.1 Time-dependent ionization modelling

Here we explicitly describe the mathematical formalism for numerical calculations on the time-dependent ionization of dynamically evolving gas exposed to a radiation field evolving in time, as motivated in Chapter 3, Section 3.6.2.

The set of time-dependent ionization equations (Chapter 3, equation (3.6.28)) may be compactly written as

$$\partial_t \boldsymbol{n} = \mathbf{A}\boldsymbol{n}, \tag{1.1.1}$$

where $\boldsymbol{n}$ is a length $N+1$ vector that specifies the ionic number densities for an element with N electrons, and $\mathbf{A}$ is the $(N+1) \times (N+1)$ matrix containing the photoionization, collisional ionization, and recombination rate coefficients, derived assuming some temperature, incident radiation field and electron number density, n_e. These equations are closed by the condition that

$$\sum_{i=0}^{N} n_{M_i} = n_{\text{total}}, \tag{1.1.2}$$

where n_{total} can be related to n_e in a highly ionized plasma, assuming solar metallicity and abundances, by

$$n_{\text{total}} \approx \frac{2X}{1+X} A_Z n_e. \tag{1.1.3}$$

Here A_Z is the absolute elemental abundance relative to hydrogen, and X is the mass fraction in hydrogen. In equilibrium, $\partial_t \boldsymbol{n} = 0$, and we find that

$$\frac{n_{M_{i+1}}}{n_{M_i}} = \frac{\beta_{M_i} n_e + \Gamma_{M_i}}{n_e \alpha_{M_i}} \equiv a_{M_i}. \tag{1.1.4}$$

It is then straightforward to show that we can write any n_{M_i} for $i > 0$ in terms of n_{M_0} via

$$n_{M_i} = \left(\prod_{j=0}^{i-1} a_{M_j} \right) n_{M_0}, \tag{1.1.5}$$

where the scaling factors a_{M_j} are defined above. Therefore

$$\begin{aligned} n_{\text{total}} = n_{M_0} + \sum_{i=1}^{N} \left(\prod_{j=0}^{i-1} a_{M_j} \right) n_{M_0} = \frac{2X}{1+X} A_Z n_e \\ \implies n_{M_0} = \frac{2X A_Z n_e}{(1+X)\left(1 + \sum_{i=1}^{N} \left(\prod_{j=0}^{i-1} a_j \right)\right)}. \end{aligned} \tag{1.1.6}$$

Eliminating n_{M_0}, we finally arrive at an expression that defines the equilibrium number density of some ion n_{M_i} as

$$n_{M_i,\text{eq}} = \prod_{j=0}^{i-1} a_{M_j} \frac{2X A_Z n_e}{(1+X)\left(1 + \sum_{k=1}^{N} \left(\prod_{j=0}^{k-1} a_{M_j} \right)\right)}, \tag{1.1.7}$$

for $i > 0$.

We first calculate the equilibrium set $n_{M_i,\text{eq}}$ using equation (1.1.7), assuming $n_e = 10^3$ cm^{-3} (approximately that inferred from the analysis in Chapter 3, Section 3.5.1) and $X = 0.28$. To examine the effect of changing gas density, we then perturb these values and n_e by a factor $(1 + \delta)$. For changes in the incident ionizing flux, we perturb the photoionization rates, Γ_{M_i}, in an identical fashion. The time-dependent evolution in the number densities, n, of H I, O IV, O V, O VI, Ne VIII, and Mg X are then calculated by numerically solving equation (1.1.1), set to initially contain the perturbed $\boldsymbol{n}$ and/or $\mathbf{A}$ (note that both $\boldsymbol{n}$ and $\mathbf{A}$ change if n_e changes). For $\mathbf{A}$, we additionally assume a temperature $T = 10^4$ K, and illumina-

tion by the 'UV peak' SED at a distance of 2.3 kpc.

In the case of H I, we can obtain an analytical solution to the rate of change in the neutral fraction $x \equiv n_{\mathrm{H\,I}}/n_{\mathrm{H}}$, determined by the hydrogen photoionization rate Γ, collisional ionization rate coefficient β, and recombination rate coefficient α, as well as the electron number density n_e, according to

$$\frac{\mathrm{d}x}{\mathrm{d}t} = -\left(\Gamma + \beta n_e\right) + \alpha n_e (1 - x). \tag{1.1.8}$$

For a pure hydrogen gas ($X = 1$), we can express the electron number density as $n_e = (1 - x)n_{\mathrm{H}}$ and write $\mathrm{d}x/\mathrm{d}t$ in the form of a Riccati equation:

$$\frac{\mathrm{d}x}{\mathrm{d}t} = Rx^2 + Qx + P, \tag{1.1.9}$$

$$R \equiv (\beta + \alpha)n_{\mathrm{H}}, \tag{1.1.10}$$

$$Q \equiv -\left(\Gamma + \alpha n_{\mathrm{H}} + R\right), \tag{1.1.11}$$

$$P \equiv \alpha n_{\mathrm{H}} \tag{1.1.12}$$

(Altay & Theuns, 2013). Assuming that P, Q and R are all constant, the time-dependent solution can then be found by separation of variables:

$$x(t) = x_- + (x_0 - x_-)\,\frac{(x_+ - x_-)\,F}{(x_+ - x_0) + (x_0 - x_-)\,F}, \tag{1.1.13}$$

where x_+ and x_- are the roots of the quadratic term in equation (A9), $x_0 \equiv x(t = 0)$ is the initial value, and

$$F(t) \equiv \exp\left(\frac{-\left(x_+ - x_-\right)t}{t_{\mathrm{rec}}}\right) \tag{1.1.14}$$

$$t_{\mathrm{rec}} \equiv \frac{1}{(\alpha + \beta)n_{\mathrm{H}}}. \tag{1.1.15}$$

We identify t_{rec} as the recombination time-scale. It can be shown that x_- represents the physical equilibrium solution (Altay & Theuns, 2013), so for the density changes described, both $x(t = \infty)$ and $x(t = 0)$ take the form of x_-, but with a different value of n_{H}. The same applies for incident ionizing flux changes, but with

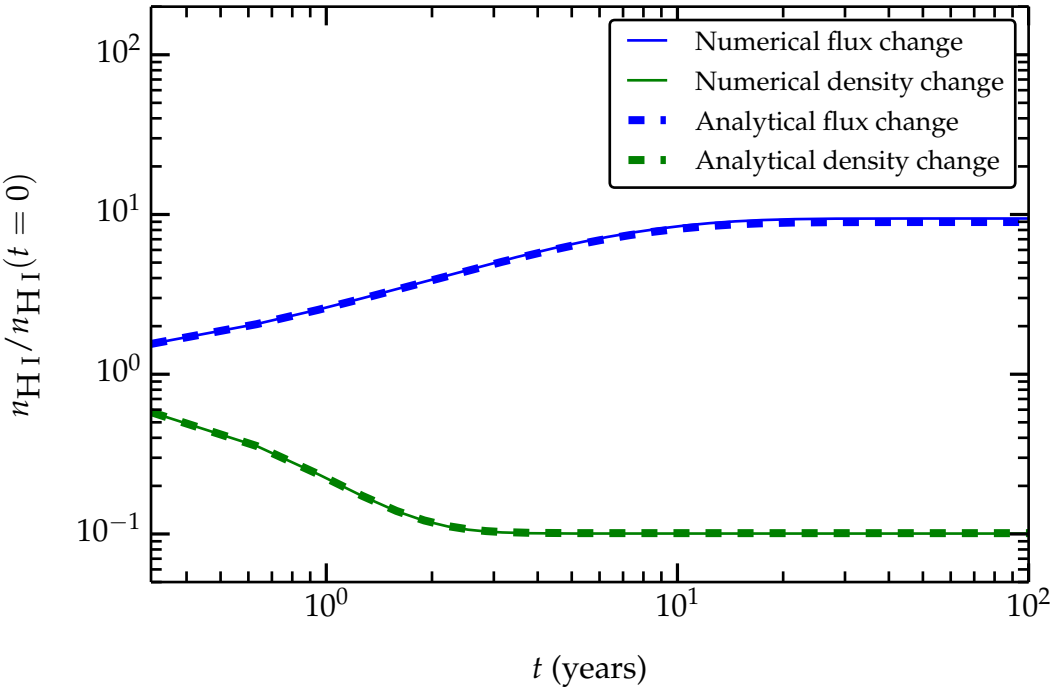

Figure A.1: Non-equilibrium evolution in $n_{\mathrm{H\,I}}$ following a step-function, order of magnitude change in flux (blue) and density (green) in a pure hydrogen gas. Numerical results are shown with solid lines, and analytical results are shown with dashed lines. The results are normalised with respect to the starting value at $t = 0$. We find an excellent agreement between the numerical and analytical calculations.

different values of Γ instead.

First, to check the validity of our numerical calculations, we compare the numerically calculated non-equilibrium evolution in the number density of H I for a pure hydrogen gas ($X = 1$) with that computed from equation (1.1.13). We do so for an order of magnitude step-function decrease in both the incident ionizing flux and the gas density separately. The results of this comparison are shown in Figure A.1. We find an excellent agreement between these two calculations, which confirms that our numerical results are robust. Full numerical results for the evolution in n following a step-function change in density are plotted in Figure A.2 relative to the starting values $n(t = 0)$, for a range of values of $(1 + \delta)$. A similar calculation for changes in the incident ionizing flux (described in Section 3.6.2) is presented in Figure A.3. We find that the numerical results in the latter agree well with the typical e-folding time-scales derived in Chapter 3, equation (3.6.29).

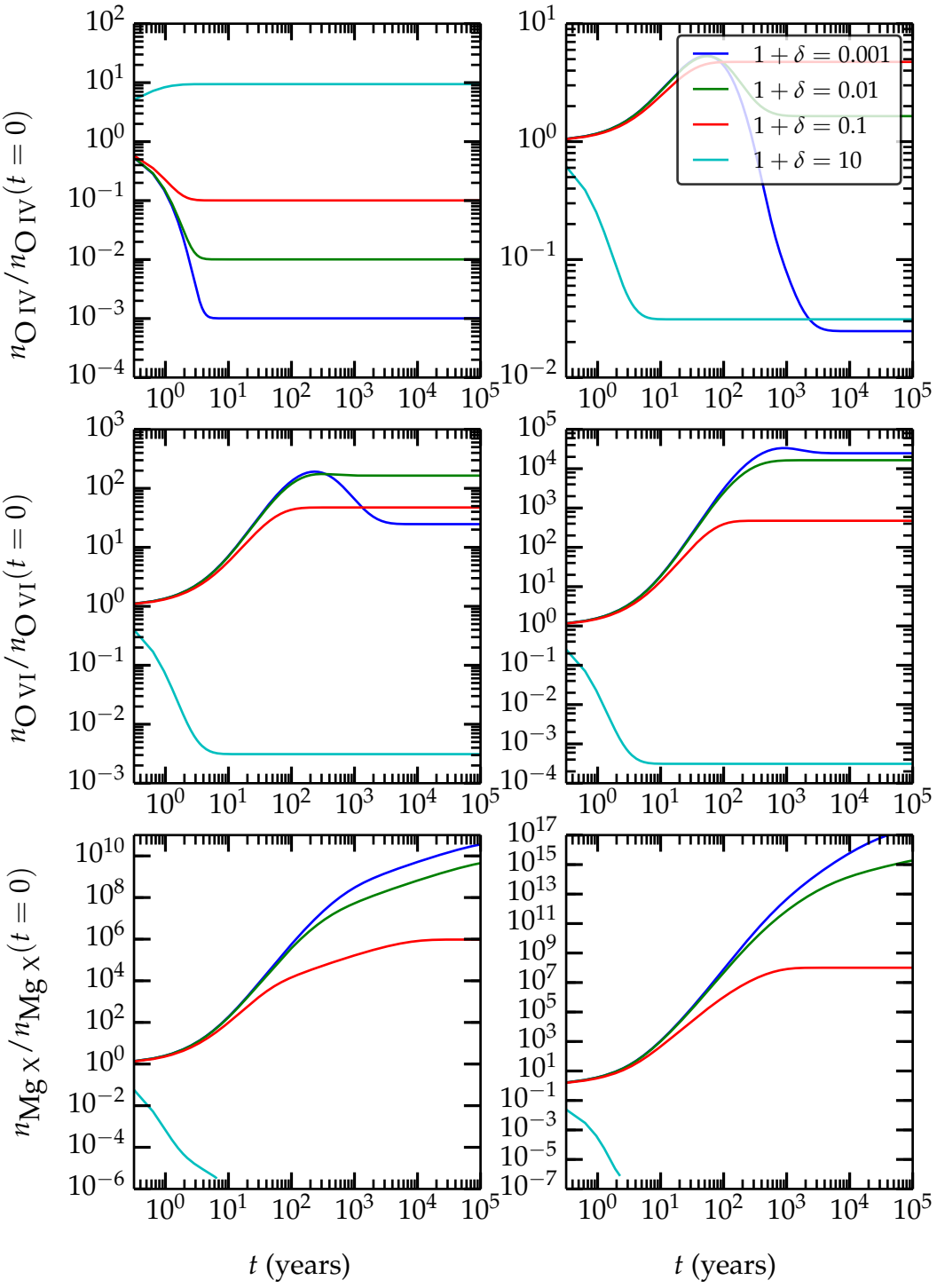

Figure A.2: Non-equilibrium evolution in the number density, n, of ions H I, O IV, O V, O VI, Ne VIII and Mg X relative to their starting values, following a step function change in density given by $n(t > 0) = (1 + \delta)n(t = 0)$. We set n at $t = 0$ to the equilibrium values for a temperature $T = 10^4$ K, a distance of 2.3 kpc from the QSO modelled by the 'UV peak' SED, and with $n_e = 10^3$ cm^{-3}. Numerical results for a range of $(1 + \delta)$ are presented with different coloured lines. Time-scales for a restored equilibrium are > 100 years for a drop in gas density of more than an order of magnitude, with the exception of H I. Time-scales are orders of magnitude shorter for an increase in gas density.

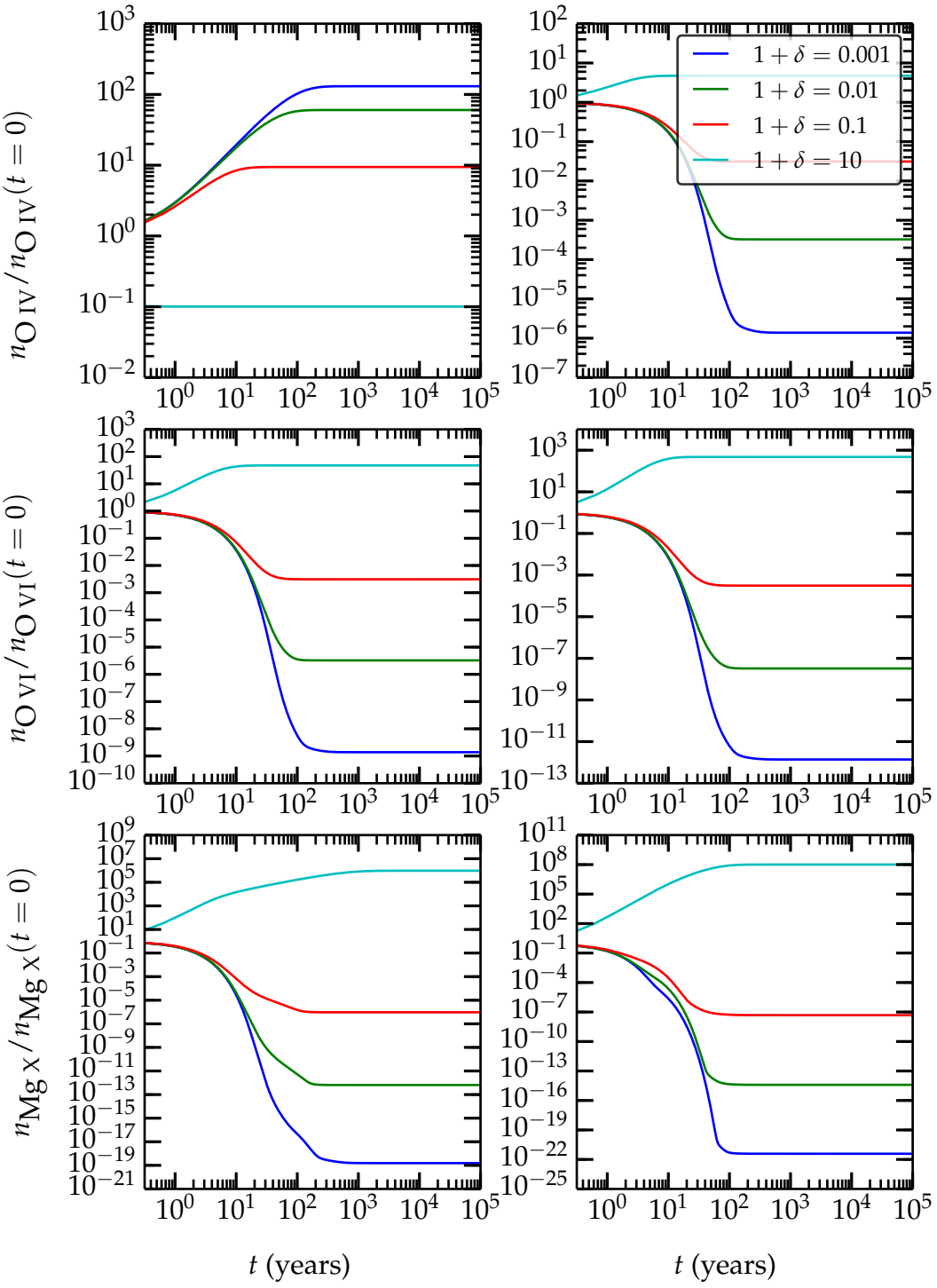

Figure A.3: Non-equilibrium evolution in the number density, n, of ions H I, O IV, O V, O VI, Ne VIII and Mg X relative to their starting values, following a step function change in incident flux given by $\Gamma(t > 0) = (1 + \delta)\Gamma(t = 0)$. We set Γ at $t = 0$ to the photoionization rate at a distance of 2.3 kpc from the QSO modelled by the 'UV peak' SED. Numerical results for a range of $(1 + \delta)$ are presented with different coloured lines. Time-scales for a restored equilibrium are typically < 100 years. These results agree well with the typical e-folding time-scales derived in equation (3.6.29).

Appendix B

B.1 Data tables

Here we present a snapshot of the tables containing data pertaining to the catalogue of quasar (QSO) absorption-line systems and the spectroscopic catalogue of galaxies around QSO sight-lines, described in Chapters 2 and 4 respectively. The full catalogues will be available in electronic format following future publication in Monthly Notices of the Royal Astronomical Society, or can be obtained upon request from the author. We use this space purely for illustration purposes.

In Table B.1, we provide a snapshot of the catalogue of QSO absorption-line systems. This catalogue contains the best-fit Voigt parameters for absorption systems attributable to the identified ions in just those QSOs analysed by the author and collaborators, namely, LBQS J0107−0235A/B, LBQS J0107−0232, FBQS J0209−0438, HE1003+0149, LBQS J1019+0147, SDSS J135726.27+043541.4, LBQS J1435−0134, and FBQS J2218+0052. All other catalogues of absorption-line systems for QSOs listed in Chapter 2, Table 2.2, are documented in Danforth et al. (2014), and can be obtained from the MAST archive.[1]

In Table B.2, we provide a snapshot of the spectroscopic catalogue of galaxies around QSO sight-lines described in Chapter 4, excluding photometric measurements available in a range of different passbands. The full catalogue contains just the data obtained and/or analysed by the author and collaborators. All other galaxy data can be obtained from the various online sources detailed in Chapter 4.

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

Table B.1: Catalogue of QSO absorption-line systems

ion	z	b (km s^{-1})	Δb (km s^{-1})	log(N/cm^{-2})	Δ log(N/cm^{-2})	label	QSO
Fe II	0.0000	18.2	3.8	15.02	0.30	a	Q0209
C IV	0.0001	32.3	3.0	13.99	0.03	a	Q0209
C IV	-0.0002	29.1	20.3	13.11	0.19	a	Q0209
C IV	-0.0006	22.4	5.3	13.33	0.06	a	Q0209
C II	-0.0000	37.3	3.7	15.46	0.18	b	Q0209
C II	-0.0003	15.9	2.8	14.29	0.05	b	Q0209
C II	-0.0005	16.3	4.8	13.85	0.05	b	Q0209
C II	-0.0007	11.8	4.3	13.77	0.06	b	Q0209
C II*	0.0001	22.4	2.7	13.98	0.03	b	Q0209
Al II	0.0000	17.4	14.6	14.56	2.61	c	Q0209
Al II	0.0002	57.8	39.6	12.70	0.46	b	Q0209
Al II	-0.0002	15.9	11.1	12.80	0.18	b	Q0209
Al II	-0.0004	27.2	4.5	13.88	0.22	b	Q0209
Si II	0.0001	18.0	5.4	15.11	0.58	a	Q0209
Si II	-0.0001	15.9	17.5	13.57	0.46	a	Q0209
Si II	-0.0003	16.4	5.3	13.58	0.06	a	Q0209
Mg II	-0.0001	34.1	18.2	14.37	0.78	a	Q0209
Mg II	-0.0006	102.5	127.6	13.11	0.56	a	Q0209
H I	0.0779	14.4	2.8	13.69	0.09	b	Q0209
H I	0.0850	40.7	23.5	12.86	0.16	b	Q0209
H I	0.0846	22.6	7.0	13.14	0.07	b	Q0209
H I	0.1015	28.0	16.6	13.69	0.16	b	Q0209
H I	0.1013	18.8	17.9	13.46	0.24	b	Q0209
H I	0.1135	28.5	9.2	13.01	0.08	b	Q0209
H I	0.1213	43.9	3.2	13.90	0.02	b	Q0209
H I	0.1290	30.6	3.7	13.61	0.04	b	Q0209
H I	0.1421	41.1	1.8	14.25	0.02	b	Q0209
H I	0.1361	41.2	5.3	14.33	0.08	b	Q0209
H I	0.1359	38.6	14.4	13.69	0.32	b	Q0209
H I	0.1367	35.0	11.3	13.13	0.09	b	Q0209
H I	0.1597	26.3	9.6	13.01	0.09	b	Q0209
H I	0.1592	52.1	6.0	13.57	0.04	b	Q0209
H I	0.1619	21.3	35.4	13.10	1.41	c	Q0209
H I	0.1617	21.5	31.5	13.60	0.73	c	Q0209
H I	0.1615	22.7	56.1	13.42	1.24	b	Q0209
H I	0.1614	22.5	128.4	13.08	5.54	c	Q0209
H I	0.1610	25.7	35.7	14.06	2.54	c	Q0209
H I	0.1605	18.9	21.9	12.74	0.61	c	Q0209
H I	0.1744	27.4	5.6	12.99	0.05	b	Q0209
H I	0.1885	34.4	9.7	12.91	0.08	b	Q0209
H I	0.1981	28.0	3.4	13.22	0.03	b	Q0209
H I	0.2108	12.5	5.0	13.03	0.06	b	Q0209
H I	0.2076	25.7	5.6	13.14	0.05	b	Q0209
H I	0.2079	16.6	7.8	12.89	0.09	b	Q0209
H I	0.2086	20.5	10.2	13.28	0.45	b	Q0209
H I	0.2087	30.1	49.3	12.93	1.04	b	Q0209
H I	0.2178	48.5	8.1	13.24	0.05	b	Q0209
...	...	...	...	...	...	...	...

Table B.2: Spectroscopic catalogue of galaxies around QSO sight-lines

ID	α (deg)	δ (deg)	z	z label	spec. type	field
100134222	151.38993	1.41244	0.1233	a	SF	J1005+0134
100134638	151.43361	1.41204	0.7285	a	non-SF	J1005+0134
100136361	151.43515	1.41681	0.0003	a	none	J1005+0134
100136507	151.38073	1.41806	0.4136	a	SF	J1005+0134
100137150	151.41892	1.42058	0.8161	a	SF	J1005+0134
100137540	151.41231	1.42246	0.3828	a	SF	J1005+0134
100137663	151.45334	1.42228	0.0001	a	none	J1005+0134
100137939	151.46291	1.41920	0.4152	a	non-SF	J1005+0134
100138520	151.43410	1.42484	0.4213	a	SF	J1005+0134
100138951	151.39180	1.42591	0.0000	a	none	J1005+0134
100138968	151.48630	1.42438	0.5038	a	SF	J1005+0134
100139199	151.48095	1.42752	0.0002	a	none	J1005+0134
100139351	151.44041	1.42831	0.4158	a	SF	J1005+0134
100139579	151.40287	1.42801	0.0005	a	none	J1005+0134
100140371	151.46029	1.43139	-1.0000	c	none	J1005+0134
100140384	151.49829	1.43059	0.7278	a	SF	J1005+0134
100140525	151.37591	1.43247	0.3112	a	SF	J1005+0134
100140686	151.41957	1.43282	1.0906	a	none	J1005+0134
100140951	151.38717	1.43422	-1.0000	c	none	J1005+0134
100142272	151.44420	1.43629	-1.0000	c	none	J1005+0134
100143800	151.49250	1.44303	0.4136	a	SF	J1005+0134
100144427	151.23561	1.44324	0.4211	a	non-SF	J1005+0134
100145085	151.45159	1.44155	0.1990	a	SF	J1005+0134
100146086	151.47729	1.44903	0.2487	a	SF	J1005+0134
100147290	151.23951	1.45179	-1.0000	c	none	J1005+0134
100148644	151.48099	1.45468	0.4155	b	non-SF	J1005+0134
100149364	151.24412	1.45703	-1.0000	c	none	J1005+0134
100149646	151.45246	1.45286	0.8925	a	SF	J1005+0134
100151150	151.20591	1.46146	-1.0000	c	none	J1005+0134
100151155	151.45455	1.45846	0.9232	b	SF	J1005+0134
100151562	151.46457	1.46082	0.1251	a	SF	J1005+0134
100151728	151.37910	1.46260	0.6173	a	SF	J1005+0134
100152365	151.28212	1.46270	0.4033	a	SF	J1005+0134
100152404	151.24884	1.46421	0.6786	b	non-SF	J1005+0134
100152653	151.29703	1.46431	-1.0000	c	none	J1005+0134
100152922	151.46659	1.46281	-1.0000	c	none	J1005+0134
100152956	151.46713	1.46496	-1.0000	c	none	J1005+0134
100154854	151.21854	1.46888	0.8429	a	SF	J1005+0134
100154976	151.25714	1.46923	0.5952	a	SF	J1005+0134
100155273	151.46619	1.46731	0.3838	a	SF	J1005+0134
100155621	151.27899	1.46669	0.0962	b	SF	J1005+0134
100155669	151.38522	1.46709	0.5009	a	non-SF	J1005+0134
100156098	151.45119	1.47169	0.1739	a	SF	J1005+0134
100156302	151.49415	1.47138	0.0683	b	SF	J1005+0134
100156333	151.47185	1.47195	-1.0000	c	none	J1005+0134
100156818	151.27301	1.47474	0.7256	a	SF	J1005+0134
100157199	151.26538	1.47336	0.6038	a	SF	J1005+0134
...	...	...	...	...	...	...

Bibliography

Aalto S., Muller S., Sakamoto K., Gallagher J. S., Martín S., Costagliola F., 2012, A&A, 546, A68

Abazajian K. N. et al., 2009, ApJS, 182, 543

Abraham R. G. et al., 2004, AJ, 127, 2455

Adelberger K. L., Steidel C. C., Shapley A. E., Pettini M., 2003, ApJ, 584, 45

Aguirre A., Schaye J., Theuns T., 2002, ApJ, 576, 1

Ahn C. P. et al., 2012, ApJS, 203, 21

Alpher R. A., Bethe H., Gamow G., 1948, Phys. Rev., 73, 803

Altay G., Theuns T., 2013, MNRAS, 434, 748

Altay G., Theuns T., Schaye J., Booth C. M., Dalla Vecchia C., 2013, MNRAS, 436, 2689

Arav N., Korista K. T., de Kool M., Junkkarinen V. T., Begelman M. C., 1999, ApJ, 516, 27

Arav N., Moe M., Costantini E., Korista K. T., Benn C., Ellison S., 2008, ApJ, 681, 954

Arav N., Borguet B., Chamberlain C., Edmonds D., Danforth C., Danforth C., 2013, MNRAS, 436, 3286

Arav N. et al., 2007, ApJ, 658, 829

Arav N. et al., 2012, A&A, 544, A33

Badnell N. R., 2006, ApJS, 167, 334

Bahcall J. N., Wolf R. A., 1968, ApJ, 152, 701

Balashev S. A., Petitjean P., Ivanchik A. V., Ledoux C., Srianand R., Noterdaeme P., Varshalovich D. A., 2011, MNRAS, 418, 357

Baldry I. K. et al., 2012, MNRAS, 421, 621

Baldwin J. A., Phillips M. M., Terlevich R., 1981, PASP, 93, 5

Balogh M. L., Pearce F. R., Bower R. G., Kay S. T., 2001, MNRAS, 326, 1228

Barbieri C. V., Fraternali F., Oosterloo T., Bertin G., Boomsma R., Sancisi R., 2005, A&A, 439, 947

Barlow T. A., Sargent W. L. W., 1997, AJ, 113, 136

Baugh C. M., Lacey C. G., Frenk C. S., Granato G. L., Silva L., Bressan A., Benson A. J., Cole S., 2005, MNRAS, 356, 1191

Benson A. J., Bower R. G., Frenk C. S., Lacey C. G., Baugh C. M., Cole S., 2003, ApJ, 599, 38

Bertin E., 2006, Astronomical Society of the Pacific Conference Series, 351, 112

Bertin E., Arnouts S., 1996, A&AS, 117, 393

Bird S., Vogelsberger M., Sijacki D., Zaldarriaga M., Springel V., Hernquist L., 2013, MNRAS, 429, 3341

Birnboim Y., Dekel A., 2003, MNRAS, 345, 349

Boomsma R., Oosterloo T. A., Fraternali F., van der Hulst J. M., Sancisi R., 2005, A&A, 431, 65

Borguet B. C. J., Edmonds D., Arav N., Dunn J., Kriss G. A., 2012, ApJ, 751, 107

Borkowski K. J., Balbus S. A., Fristrom C. C., 1990, ApJ, 355, 501

Bouché N., Murphy M. T., Kacprzak G. G., Péroux C., Contini T., Martin C. L., Dessauges-Zavadsky M., 2013, Science, 341, 50

Bower R. G., Benson A. J., Malbon R., Helly J. C., Frenk C. S., Baugh C. M., Cole S., Lacey C. G., 2006, MNRAS, 370, 645

Brandt W. N., Laor A., Wills B. J., 2000, ApJ, 528, 637

Bregman J. N., 1980, ApJ, 236, 577

Brinchmann J., Charlot S., White S. D. M., Tremonti C., Kauffmann G., Heckman T., Brinkmann J., 2004, MNRAS, 351, 1151

Brüns C. et al., 2005, A&A, 432, 45

Bruzual G., Charlot S., 2003, MNRAS, 344, 1000

Burbidge E. M., Lynds C. R., Burbidge G. R., 1966, ApJ, 144, 447

Calzetti D., 2001, PASP, 113, 1449

Cantalupo S., 2010, MNRAS, 403, L16

Cantalupo S., Lilly S. J., Haehnelt M. G., 2012, MNRAS, 425, 1992

Cardelli J. A., Clayton G. C., Mathis J. S., 1989, ApJ, 345, 245

Carroll S. M., 2001, Living Reviews in Relativity, 4, 1

Carswell R. F., Whelan J. A. J., Smith M. G., Boksenberg A., Tytler D., 1982, MNRAS, 198, 91

Cen R., Ostriker J. P., 1999, ApJ, 514, 1

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

Chaves T. A., Irwin J. A., 2001, ApJ, 557, 646

Chen H. W., Mulchaey J. S., 2009, ApJ, 701, 1219

Cicone C. et al., 2014, A&A, 562, A21

Ciotti L., D'Ercole A., Pellegrini S., Renzini A., 1991, ApJ, 376, 380

Clowe D., Bradač M., Gonzalez A. H., Markevitch M., Randall S. W., Jones C., Zaritsky D., 2006, ApJ, 648, L109

Colless M. et al., 2001, MNRAS, 328, 1039

Costa T., Sijacki D., Haehnelt M. G., 2015, MNRAS, 448, L30

Crain R. A., McCarthy I. G., Schaye J., Theuns T., Frenk C. S., 2013, MNRAS, 432, 3005

Crain R. A. et al., 2009, MNRAS, 399, 1773

Creasey P., Theuns T., Bower R. G., 2013, MNRAS, 429, 1922

Crenshaw D. M., Kraemer S. B., George I. M., 2003, ARA&A, 41, 117

Crighton N. H. M., Morris S. L., Bechtold J., Crain R. A., Jannuzi B. T., Shone A., Theuns T., 2010, MNRAS, 402, 1273

Crighton N. H. M., Hennawi J. F., Prochaska J. X., 2013, ApJ, 776, L18

Cuillandre J. C. J. et al., 2012, Proc. SPIE, 8448, 84480M

Cushing M. C., Vacca W. D., Rayner J. T., 2004, PASP, 116, 362

Dalla Vecchia C., Schaye J., 2012, MNRAS, 426, 140

Danforth C. W. et al., 2014, preprint (arXiv:1402.2655)

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

Davé R., Oppenheimer B. D., Katz N., Kollmeier J. A., Weinberg D. H., 2010, MNRAS, 408, 2051

Davé R., Finlator K., Oppenheimer B. D., 2012, MNRAS, 421, 98

Davé R. et al., 2001, ApJ, 552, 473

Davis M., Peebles P. J. E., 1983, ApJ, 267, 465

Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, ApJ, 292, 371

de Kool M., Begelman M. C., 1995, ApJ, 455, 448

de Kool M., Korista K. T., Arav N., 2002, ApJ, 580, 54

Debes J. et al., 2015, Cosmic Origins Spectrograph Instrument Handbook. Version 7.0 (Baltimore: STScI)

Dekel A., Birnboim Y., 2006, MNRAS, 368, 2

Dekel A., Sari R., Ceverino D., 2009, ApJ, 703, 785

Dekel A., Zolotov A., Tweed D., Cacciato M., Ceverino D., Primack J. R., 2013, MNRAS, 435, 999

Di Gesu L. et al., 2013, A&A, 556, A94

Di Matteo T., Springel V., Hernquist L., 2005, Nature, 433, 604

Dicke R. H., Peebles P. J. E., Roll P. G., Wilkinson D. T., 1965, ApJ, 142, 414

Dickey J. M., 1990, ARA&A, 28, 215

D'Odorico V., Cristiani S., Romano D., Granato G. L., Danese L., 2004, MNRAS, 351, 976

Dolag K., Borgani S., Murante G., Springel V., 2009, MNRAS, 399, 497

Done C., Davis S. W., Jin C., Blaes O., Ward M., 2012, MNRAS, 420, 1848

Draine B. T., 2011, Physics of the Interstellar and Intergalactic Medium. Princeton Univ. Press, Princeton

Driver S. P. et al., 2011, MNRAS, 413, 971

Edmonds D. et al., 2011, ApJ, 739, 7

Einstein A., 1916, Ann. Phys., 354, 769

Faucher-Giguère C. A., Quataert E., Murray N., 2012, MNRAS, 420, 1347

Feldmann R., 2013, MNRAS, 433, 1910

Ferland G. J. et al., 2013, RMxAA, 49, 137

Feruglio C., Maiolino R., Piconcelli E., Menci N., Aussel H., Lamastra A., Fiore F., 2010, A&A, 518, L155

Finlator K., Davé R., 2008, MNRAS, 385, 2181

Finn C. W. et al., 2014, MNRAS, 440, 3317

Fischera J., Fischera J., Dopita M., Dopita M., 2005, ApJ, 619, 340

Foltz C. B., Weymann R. J., Peterson B. M., Sun L., Malkan M. A., Chaffee F. H. J., 1986, ApJ, 307, 504

Forbes D. A., Kroupa P., 2011, PASA, 28, 77

Ford A. B., Oppenheimer B. D., Davé R., Katz N., Kollmeier J. A., Weinberg D. H., 2013, MNRAS, 432, 89

Ford H. C. et al., 1998, Proc. SPIE, 3356, 234

Forman W., Jones C., Tucker W., 1985, ApJ, 293, 102

Fraternali F., Binney J. J., 2006, MNRAS, 366, 449

Fraternali F., Binney J. J., 2008, MNRAS, 386, 935

Fraternali F., van Moorsel G., Sancisi R., Oosterloo T., 2002, AJ, 123, 3124

Fraternali F., Marasco A., Marinacci F., Binney J., 2013, ApJ, 764, L21

Friedmann A., 1922, Z. Phys., 10, 377

Fukugita M., Peebles P. J. E., 2004, ApJ, 616, 643

Fukugita M., Peebles P. J. E., 2006, ApJ, 639, 590

Fukugita M., Hogan C. J., Peebles P. J. E., 1998, ApJ, 503, 518

Fumagalli M., Prochaska J. X., Kasen D., Dekel A., Ceverino D., Primack J. R., 2011, MNRAS, 418, 1796

Fumagalli M., O'Meara J. M., Prochaska J. X., Worseck G., 2013, ApJ, 775, 78

Gabel J. R., Arav N., Kim T. S., 2006, ApJ, 646, 742

Gabel J. R. et al., 2005, ApJ, 623, 85

Ganguly R., Eracleous M., Charlton J. C., Churchill C. W., 1999, AJ, 117, 2594

Ganguly R. et al., 2013, MNRAS, 435, 1233

Gehrels N., 1986, ApJ, 303, 336

Ghavamian P. et al., 2009, Preliminary Characterization of the Post-Launch Line Spread Function of COS, STScI ISR COS 2009-01, STScI, Baltimore

Gnedin N. Y., Kravtsov A. V., 2010, ApJ, 714, 287

Gorenstein P., 1975, ApJ, 198, 95

Green J. C. et al., 2012, ApJ, 744, 60

Groth E. J., Peebles P. J. E., 1977, ApJ, 217, 385

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

Guzzo L. et al., 2014, A&A, 566, A108

Haardt F., Madau P., 2001, in D.M. Neumann, J.T.V. Tran, eds, Clusters of Galaxies and the High Redshift Universe Observed in X-rays

Hall P. B., Anosov K., White R. L., Brandt W. N., Gregg M. D., Gibson R. R., Becker R. H., Schneider D. P., 2011, MNRAS, 411, 2653

Hamann F., Ferland G., 1993, ApJ, 418, 11

Hamann F., Ferland G., 1999, ARA&A, 37, 487

Hamann F., Sabra B., 2004, Astronomical Society of the Pacific Conference Series, 311, 203

Hamann F., Barlow T. A., Beaver E. A., Burbidge E. M., Cohen R. D., Junkkarinen V., Lyons R., 1995, ApJ, 443, 606

Hamann F., Barlow T. A., Junkkarinen V., Burbidge E. M., 1997, ApJ, 478, 80

Hamann F., Chartas G., McGraw S., Rodríguez Hidalgo P., Shields J., Capellupo D., Charlton J., Eracleous M., 2013, MNRAS, 435, 133

Hamann F. W., Netzer H., Shields J. C., 2000, ApJ, 536, 101

Hamann F. W., Barlow T. A., Chaffee F. C., Foltz C. B., Weymann R. J., 2001, ApJ, 550, 142

Hernandez A. K., Wakker B. P., Benjamin R. A., French D., Kerp J., Lockman F. J., O'Toole S., Winkel B., 2013, ApJ, 777, 19

Hernquist L., Katz N., Weinberg D. H., Miralda-Escudé J., 1996, ApJ, 457, L51

Hockney R. W., Eastwood J. W., 1981, Computer Simulation Using Particles. McGraw-Hill, New York

Hoopes C. G., Walterbos R. A. M., Rand R. J., 1999, ApJ, 522, 669

Hopkins A. M. et al., 2003, ApJ, 599, 971

Hopkins A. M. et al., 2013, MNRAS, 430, 2047

Hopkins P. F., Elvis M., 2010, MNRAS, 401, 7

Hubble E., 1929, Proceedings of the National Academy of Science, 15, 168

Hummels C. B., Bryan G. L., Smith B. D., Turk M. J., 2013, MNRAS, 430, 1548

Jiménez-Vicente J., Mediavilla E., Muñoz J. A., Kochanek C. S., 2012, ApJ, 751, 106

Jin C., Ward M., Done C., Gelbord J., 2012, MNRAS, 420, 1825

Johnson S. D., Chen H. W., Mulchaey J. S., 2013, MNRAS, 434, 1765

Kaiser N., 1987, MNRAS, 227, 1

Kalberla P. M. W., Burton W. B., Hartmann D., Arnal E. M., Bajaja E., Morras R., Pöppel W. G. L., 2005, A&A, 440, 775

Kaspi S. et al., 2002, ApJ, 574, 643

Katz N., Weinberg D. H., Hernquist L., 1996, ApJS, 105, 19

Keeney B. A., Danforth C. W., Stocke J. T., France K., Green J. C., 2012, PASP, 124, 830

Keeney B. A., Stocke J. T., Rosenberg J. L., Danforth C. W., Ryan-Weber E. V., Shull J. M., Savage B. D., Green J. C., 2013, ApJ, 765, 27

Kennicutt R. C. J., 1992, ApJ, 388, 310

Kennicutt R. C. J., 1998a, ARA&A, 36, 189

Kennicutt R. C. J., 1998b, ApJ, 498, 541

Kereš D., Katz N., Weinberg D. H., Davé R., 2005, MNRAS, 363, 2

Kereš D., Katz N., Davé R., Fardal M., Weinberg D. H., 2009a, MNRAS, 396, 2332

Kereš D., Katz N., Fardal M., Davé R., Weinberg D. H., 2009b, MNRAS, 395, 160

Kewley L. J., Dopita M. A., Sutherland R. S., Heisler C. A., Trevena J., 2001, ApJ, 556, 121

Kewley L. J., Geller M. J., Jansen R. A., 2004, AJ, 127, 2002

King A., 2003, ApJ, 596, L27

Kinman T. D., Lamla E., Wirtanen C. A., 1966, ApJ, 146, 964

Knuth K. H., 2006, preprint (arXiv:physics/0605197)

Komatsu E. et al., 2011, ApJS, 192, 18

Kriss G. A., 2011, Improved Medium Resolution Line Spread Functions for COS FUV Spectra, STScI ISR COS 2011-01, STScI, Baltimore

Krolik J. H., Kriss G. A., 1995, ApJ, 447, 512

Kuhlen M., Krumholz M. R., Madau P., Smith B. D., Wise J., 2012, ApJ, 749, 36

Kwak K., Shelton R. L., 2010, ApJ, 719, 523

Lagos C. D. P., Cora S. A., Padilla N. D., 2008, MNRAS, 388, 587

Landy S. D., Szalay A. S., 1993, ApJ, 412, 64

Le Fevre O. et al., 2004, A&A, 417, 839

Le Fevre O. et al., 2005, A&A, 439, 845

Lee J. C. et al., 2013, MNRAS, 430, 2650

Lee K. G. et al., 2014, ApJ, 795, L12

Lehner N., Howk J. C., 2011, Science, 334, 955

Lemaître G., 1927, Ann. Soc. Sci. Bruxelles, 47, 49

Li C., White S. D. M., 2009, MNRAS, 398, 2177

Li J. T., Wang Q. D., 2013a, MNRAS, 428, 2085

Li J. T., Wang Q. D., 2013b, MNRAS, 435, 3071

Li J. T., Crain R. A., Wang Q. D., 2014, MNRAS, 440, 859

Lilly S. J., Carollo C. M., Pipino A., Renzini A., Peng Y., 2013, ApJ, 772, 119

Lynds C. R., Stockton A. N., 1966, ApJ, 144, 446

Maiolino R. et al., 2012, MNRAS, 425, L66

Marasco A., Fraternali F., Binney J. J., 2012, MNRAS, 419, 1107

Maraston C., Strömbäck G., 2011, MNRAS, 418, 2785

Marinacci F., Binney J., Fraternali F., Nipoti C., Ciotti L., Londrillo P., 2010, MNRAS, 404, 1464

Mathews W. G., 1990, ApJ, 354, 468

Mathewson D. S., Cleary M. N., Murray J. D., 1974, ApJ, 190, 291

Mathur S., Wilkes B., Elvis M., Fiore F., 1994, ApJ, 434, 493

Mathur S., Wilkes B., Elvis M., 1998, ApJ, 503, L23

Matsuoka K. et al., 2013, ApJ, 771, 64

Matthews L. D., Wood K., 2003, ApJ, 593, 721

Mazzotta P., Mazzitelli G., Colafrancesco S., Vittorio N., 1998, A&AS, 133, 403

McConnell N. J., Ma C. P., 2013, ApJ, 764, 184

McGaugh S. S., Schombert J. M., de Blok W. J. G., Zagursky M. J., 2010, ApJ, 708, L14

McPhate J. B., Siegmund O. H., Gaines G. A., Vallerga J. V., Hull J. S., 2000, Proc. SPIE, 4139, 25

Miller S. T., Veilleux S., 2003a, ApJ, 148, 383

Miller S. T., Veilleux S., 2003b, ApJ, 592, 79

Miralda-Escudé J., Cen R., Ostriker J. P., Rauch M., 1996, ApJ, 471, 582

Misawa T., Charlton J. C., Eracleous M., Ganguly R., Tytler D., Kirkman D., Suzuki N., Lubin D., 2007, ApJ, 171, 1

Misgeld I., Hilker M., 2011, MNRAS, 414, 3699

Mo H., van den Bosch F. C., White S., 2010, Galaxy Formation and Evolution. Cambridge Univ. Press

Moe M., Arav N., Bautista M. A., Korista K. T., 2009, ApJ, 706, 525

Morris S. L., Jannuzi B. T., 2006, MNRAS, 367, 1261

Morris S. L., Weymann R. J., Foltz C. B., Turnshek D. A., Shectman S., Price C., Boroson T. A., 1986, ApJ, 310, 40

Morton D. C., 2003, ApJS, 149, 205

Morton D. C., York D. G., Jenkins E. B., 1988, ApJS, 68, 449

Moustakas J., Kennicutt R. C. J., Tremonti C. A., 2006, ApJ, 642, 775

Muzahid S., Srianand R., Savage B. D., Narayanan A., Mohan V., Dewangan G. C., 2012, MNRAS, 424, L59

Muzahid S., Srianand R., Arav N., Savage B. D., Narayanan A., 2013, MNRAS, 431, 2885

Nestor D., Hamann F., Rodríguez Hidalgo P., 2008, MNRAS, 386, 2055

Nicastro F., Fiore F., Perola G. C., Elvis M., 1999, ApJ, 512, 184

Novak G. S., Ostriker J. P., Ciotti L., 2011, ApJ, 737, 26

Oosterloo T. A., Morganti R., Sadler E. M., van der Hulst T., Serra P., 2007, A&A, 465, 787

Oppenheimer B. D., Schaye J., 2013a, MNRAS, 434, 1063

Oppenheimer B. D., Schaye J., 2013b, MNRAS, 434, 1043

Osterbrock D. E., 1989, Astrophysics of gaseous nebulae and active galactic nuclei. University Science Books, Mill Valley, CA

Osterman S. et al., 2011, Ap&SS, 335, 257

Ostriker J. P., Choi E., Ciotti L., Novak G. S., Proga D., 2010, ApJ, 722, 642

Parriott J. R., Bregman J. N., 2008, ApJ, 681, 1215

Penton S. V., Stocke J. T., Shull J. M., 2002, ApJ, 565, 720

Penzias A. A., Wilson R. W., 1965, ApJ, 142, 419

Perlmutter S. et al., 1999, ApJ, 517, 565

Petitjean P., Srianand R., 1999, A&A, 345, 73

Petitjean P., Rauch M., Carswell R. F., 1994, A&A, 291, 29

Pettini M., Cooke R., 2012, MNRAS, 425, 2477

Planck Collaboration et al., 2014, A&A, 571, A16

Press W. H., Schechter P., 1974, ApJ, 187, 425

Prochaska J. X., Weiner B., Chen H. W., Cooksey K. L., Mulchaey J. S., 2011a, ApJ, 193, 28

Prochaska J. X., Weiner B., Chen H. W., Mulchaey J., Cooksey K., 2011b, ApJ, 740, 91

Putman M. E., Peek J. E. G., Joung M. R., 2012, ARA&A, 50, 491

Rees M. J., Ostriker J. P., 1977, MNRAS, 179, 541

Richter P., 2012, ApJ, 750, 165

Riess A. G. et al., 1998, AJ, 116, 1009

Risaliti G., Elvis M., 2010, A&A, 516, A89

Robertson H. P., 1935, ApJ, 82, 284

Rossa J., Dettmar R. J., 2000, A&A, 359, 433

Rossa J., Dettmar R. J., 2003a, A&A, 406, 493

Rossa J., Dettmar R. J., 2003b, A&A, 406, 505

Rubin V. C., Ford W. K. J., 1970, ApJ, 159, 379

Rubin V. C., Ford W. K. J., Thonnard N., 1980, ApJ, 238, 471

Sargent W. L. W., Boksenberg A., Young P., 1982, ApJ, 252, 54

Savage B. D., Kim T. S., Wakker B. P., Keeney B., Shull J. M., Stocke J. T., Green J. C., 2014, ApJS, 212, 8

Sawala T. et al., 2015, MNRAS, 448, 2941

Scannapieco C. et al., 2012, MNRAS, 2970

Scannapieco E., Oh S. P., 2004, ApJ, 608, 62

Schaye J., 2001, ApJ, 559, 507

Schaye J. et al., 2010, MNRAS, 402, 1536

Schaye J. et al., 2015, MNRAS, 446, 521

Scheuer P. A. G., 1965, Nature, 207, 963

Schmidt P., Józsa G. I. G., Gentile G., Oh S. H., Schuberth Y., Ben Bekhti N., Winkel B., Klein U., 2014, A&A, 561, A28

Seaton M. J., 1979, MNRAS, 187, 73P

Sembach K. R. et al., 2003, ApJS, 146, 165

Sharma M., Nath B. B., Chand H., 2013, MNRAS, 431, L93

Shull J. M., 2014, ApJ, 784, 142

Shull J. M., Jones J. R., Danforth C. W., Collins J. A., 2009, ApJ, 699, 754

Shull J. M., Stevans M., Danforth C. W., 2012, ApJ, 752, 162

Sijacki D., Vogelsberger M., Kereš D., Springel V., Hernquist L., 2012, MNRAS, 424, 2999

Silk J., Rees M. J., 1998, A&A, 331, L1

Simcoe R. A. et al., 2013, PASP, 125, 270

Skrutskie M. F. et al., 2006, AJ, 131, 1163

Spitzer L. J., 1956, ApJ, 124, 20

Springel V., White S. D. M., Tormen G., Kauffmann G., 2001, MNRAS, 328, 726

Srianand R., Petitjean P., 2000, A&A, 357, 414

Srianand R., Petitjean P., 2001, A&A, 373, 816

Srianand R., Shankaranarayanan S., 1999, ApJ, 518, 672

Stocke J. T., Keeney B. A., Danforth C. W., Shull J. M., Froning C. S., Green J. C., Penton S. V., Savage B. D., 2013, ApJ, 763, 148

Stocke J. T. et al., 2014, ApJ, 791, 128

Stockton A. N., Lynds C. R., 1966, ApJ, 144, 451

Storey P. J., Zeippen C. J., 2000, MNRAS, 312, 813

Stoughton C. et al., 2002, AJ, 123, 485

Strickland D. K., Heckman T. M., Colbert E. J. M., Hoopes C. G., Weaver K. A., 2004a, ApJ, 151, 193

Strickland D. K., Heckman T. M., Colbert E. J. M., Hoopes C. G., Weaver K. A., 2004b, ApJ, 606, 829

Sturm E. et al., 2011, ApJ, 733, L16

Tayal S. S., 2006, ApJS, 166, 634

Tejos N., Morris S. L., Crighton N. H. M., Theuns T., Altay G., Finn C. W., 2012, MNRAS, 425, 245

Tejos N. et al., 2014, MNRAS, 437, 2017

Telfer R. C., Kriss G. A., Zheng W., Davidsen A. F., Green R. F., 1998, ApJ, 509, 132

Tepper-García T., Richter P., Schaye J., Booth C. M., Dalla Vecchia C., Theuns T., Wiersma R. P. C., 2011, MNRAS, 413, 190

Tepper-García T., Richter P., Schaye J., Booth C. M., Dalla Vecchia C., Theuns T., 2012, MNRAS, 425, 1640

Tepper-García T., Richter P., Schaye J., 2013, MNRAS, 436, 2063

Theuns T., Leonard A., Efstathiou G., Pearce F. R., Thomas P. A., 1998, MNRAS, 301, 478

Tripp T. M., Lu L., Savage B. D., 1996, ApJS, 102, 239

Trump J. R. et al., 2006, ApJS, 165, 1

Tumlinson J., Fang T., 2005, ApJ, 623, L97

Tumlinson J. et al., 2011, Science, 334, 948

Tumlinson J. et al., 2013, ApJ, 777, 59

Turner M. L., Schaye J., Steidel C. C., Rudie G. C., Strom A. L., 2014, MNRAS, 445, 794

van de Voort F., Schaye J., Booth C. M., Haas M. R., Dalla Vecchia C., 2011, MNRAS, 414, 2458

Veilleux S., Cecil G., Bland-Hawthorn J., 2005, ARA&A, 43, 769

Verner D. A., Barthel P. D., Tytler D., 1994, A&AS, 108, 287

Véron-Cetty M. P., Joly M., Véron P., 2004, A&A, 417, 515

Vivek M., Srianand R., Petitjean P., Noterdaeme P., Mohan V., Mahabal A., Kuriakose V. C., 2012, MNRAS, 423, 2879

Vogelsberger M. et al., 2014, MNRAS, 444, 1518

Voges W. et al., 1999, A&A, 349, 389

Voronov G. S., 1997, At. Data. Nucl. Data Tables, 65, 1

Wakker B. P., van Woerden H., 1997, ARA&A, 35, 217

Walker A. G., 1935, MNRAS, 95, 263

Wang Q. D., Immler S., Walterbos R., Lauroesch J. T., Breitschwerdt D., 2001, ApJ, 555, L99

Weinberg S., 1989, Rev. Mod. Phys., 61, 1

Werk J. K., Prochaska J. X., Thom C., Tumlinson J., Tripp T. M., O'Meara J. M., Peeples M. S., 2013, ApJS, 204, 17

Werk J. K. et al., 2014, ApJ, 792, 8

Weymann R. J., Williams R. E., Peterson B. M., Turnshek D. A., 1979, ApJ, 234, 33

Weymann R. J., Morris S. L., Foltz C. B., Hewett P. C., 1991, ApJ, 373, 23

White S. D. M., Frenk C. S., 1991, ApJ, 379, 52

White S. D. M., Rees M. J., 1978, MNRAS, 183, 341

Wiersma R. P. C., Schaye J., Dalla Vecchia C., Booth C. M., Theuns T., Aguirre A., 2010, MNRAS, 409, 132

Wild V. et al., 2008, MNRAS, 388, 227

Williams R. E., Strittmatter P. A., Carswell R. F., Craine E. R., 1975, ApJ, 202, 296

Willman B., Strader J., 2012, AJ, 144, 76

Wilms J., Allen A., McCray R., 2000, ApJ, 542, 914

Woo J. H., Urry C. M., 2002, ApJ, 579, 530

Woodgate B. E. et al., 1998, PASP, 110, 1183

Young P. J., Sargent W. L. W., Boksenberg A., Carswell R. F., Whelan J. A. J., 1979, ApJ, 229, 891

Zubovas K., King A. R., 2014, MNRAS, 439, 400

Zwicky F., 1937, ApJ, 86, 217