

HI Absorption in the Intergalactic Medium

J. Xavier Prochaska

Abstract We review studies of the intergalactic medium (IGM) via HI absorption with emphasis on the decades of previous research with quasar absorption line studies. The chapter begins with a historical perspective and then offers a pedagogical description of the quantum mechanics underlying the Lyman series and Lyman continuum opacity of the HI atom. We describe the manifestation of these opacities in absorption spectroscopy and the challenges related to normalization of quasar emission. Standard measurement techniques (equivalent width, line-profile fitting) are introduced. We then introduce the N_{HI} frequency distribution $f(N_{\text{HI}})$ and efforts to constrain this distribution across cosmic time with emphasis on the optically thin Ly α forest. A discussion of optically gas (Lyman limit systems and damped Ly α systems) and its relation to the mean free path is presented. Online presentations and Python notebooks supplement this chapter with examples, a review of modern work, and thoughts on future progress. See <https://github.com/profxj/SaasFee2016> for additional resources.

1 Historical Introduction

The discovery of the intergalactic medium (IGM) was, in essence, precipitated by the discovery of quasars in 1963¹ by Schmidt (1963). It was through spectroscopy of these enigmatic, distant sources that one could resolve the absorption lines from gas – especially HI Ly α – in the foreground universe. Figure 1 shows an early example from Burbidge et al (1966) taken with the prime-focus spectrograph on the Shane 120-inch telescope at Lick Observatory. Even in these early data, one identifies apparently discrete absorption lines of Hydrogen and heavy elements establishing the

J. Xavier Prochaska

University of California, 1156 High St., Santa Cruz, CA 95064 USA e-mail: xavier@ucolick.org

¹ There were, however, unsuccessful attempts to search for extragalactic gas in 21 cm absorption (Field, 1959).

presence of diffuse yet enriched gas along the sightline. This thicket of absorption blueward of the quasar's Ly α emission is now commonly referred to as the Ly α Forest.

Spectra like these inspired the first models of the IGM as discrete absorption lines (Bahcall and Salpeter, 1966) and by inference the first physical insight. Gunn and Peterson (1965) recognized that a universe with predominantly neutral hydrogen gas should be opaque to these far-UV photons and inferred – correctly – that the gas must have a neutral fraction x_{HI} of less than 1 part in 10^5 . That is, the positive detection of flux at rest wavelengths shortward of the quasar Ly α emission line ($\lambda_{\text{rest}} < 1215\text{\AA}$), i.e. transmission through the Ly α forest, demands a highly ionized IGM. As an introduction to the material presented in this Chapter, we may offer our own rough estimate. The optical depth of HI Ly α through a $\Delta d = 100\text{ kpc}$ portion of the $z = 3$ universe at the mean hydrogen density \bar{n}_{H} is simply

$$\tau(v) = \Delta d \bar{n}_{\text{H}} x_{\text{HI}} \sigma_{\text{Ly}\alpha}(v) \quad (1)$$

with $\sigma_{\text{Ly}\alpha}$ the Ly α cross-section. We estimate the latter assuming Doppler broadening dominates with a characteristic velocity given by Hubble expansion, $\Delta v \approx H(z) * \ell \approx 30\text{ km/s}$. Taking a baryonic mass density $\rho_b = 0.0486\rho_c$ at $z = 0$ with ρ_c the critical density and taking 75% of the baryonic mass as Hydrogen, we find $\tau(v) \approx 10^6 x_{\text{HI}}$. Therefore, the positive detection of flux in the Ly α forest demands a highly ionized IGM.

The remainder of the 1960's introduced a series of fundamental papers on the astrophysics of absorption-line analysis, especially by Bachall and his collaborators. These included the discussion of fundamental diagnostics of the gas (Bahcall, 1967), the application of absorption from the fine-structure levels of heavy elements (Bab-

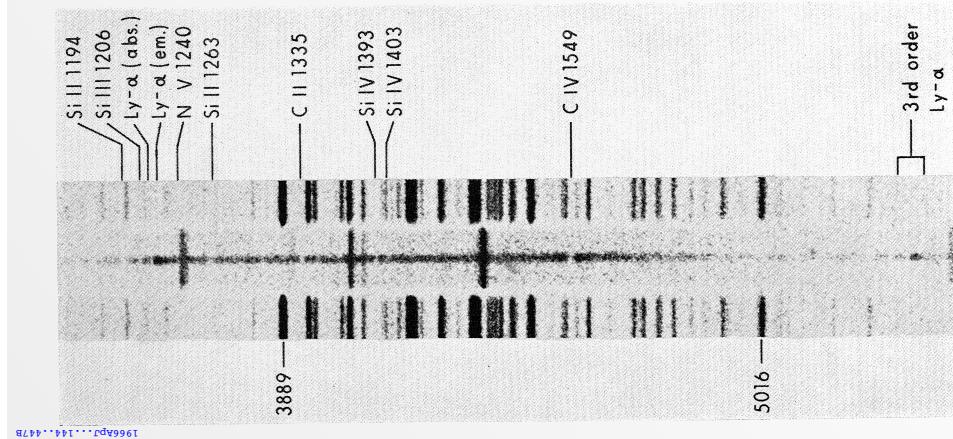


Fig. 1 Lick spectrum of 3C 191 obtained in February 1966 with the prime-focus spectrograph on the Shane 120-inch telescope at Lick Observatory. The comparison lamp spectrum shown is that of He+Ar. Taken from Burbidge et al (1966).

call and Wolf, 1968), and the assertion that the majority of heavy element absorption may be associated to the halos of galaxies (Bahcall and Peebles, 1969; Bahcall and Spitzer, 1969). In a number of respects, the theory had outpaced the observations. This held throughout the 1970's, especially for IGM studies with HI Ly α although Brown and Roberts (1973) eventually reported the first intergalactic detection of HI in 21 cm absorption.

In the early 1980's, advances in spectroscopic technology (especially the CCD detector) led to the first high-quality views of the HI Ly α forest (Figure 2; Young et al, 1979; Boksenberg et al, 1978; Sargent et al, 1980). It was evident from spectra like these that the IGM was characterized by a stochastic forest of HI absorption well-described by discrete lines. In essence, these were the first detailed views of the Ly α forest. This decade also witnessed the first surveys on gas optically thick at the HI Lyman limit (aka Lyman Limit Systems or LLSs; Tytler, 1982) and on the HI Ly α absorbers with sufficient column density to generate damped Ly α profiles (aka damped Ly α systems or DLAs; Wolfe et al, 1986), and the first empirical connections between absorption and individual galaxies (Bergeron and Boisse, 1991). The field was suddenly awash with data and theory had now fallen behind. The observers took to developing models of 'spherical' HI clouds and bull's-eye cartoons to describe the gas around galaxies. At that time, J. Ostriker was the most active theorist on the IGM, publishing a series of papers on applications of the IGM including its first phase diagram (Ostriker and Ikeuchi, 1983; Ostriker and Heisler, 1984; Bajtlik et al, 1988; Duncan et al, 1989). But one of his leading models of the day envisioned "Galaxy formation in an IGM dominated by explosion" (Ostriker and Cowie, 1981). A deeper understanding of the IGM was still to be developed.

In several respects, the 1990's witnessed the true maturation of studies on the IGM. Observationally, the HIRES spectrometer (Vogt et al, 1994) on the 10-m W.M. Keck telescope fully resolved the IGM and at terrific S/N. These spectra represent the pinnacle, analogous to Planck measurements of the CMB. The advance over even the 1980's was profound, as Figure 3 illustrates.

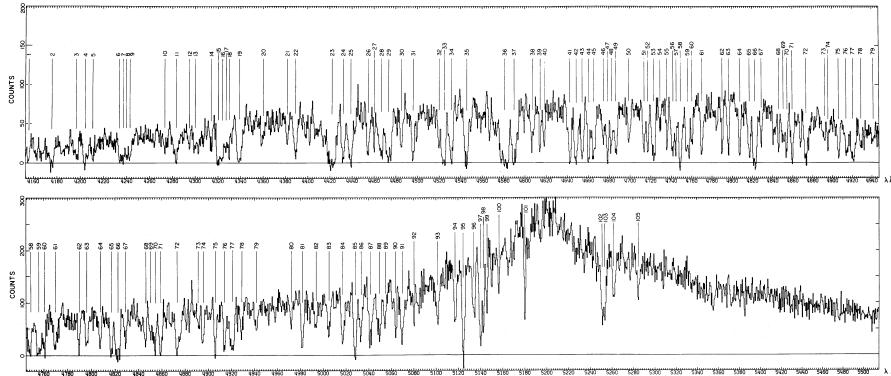


Fig. 2 HI Ly α forest spectrum of the quasar PKS 2126–158 obtained by Young et al (1979). Data like these provided the first detailed view of the discrete absorption of the Ly α forest.

A new paradigm for the IGM emerged from hydrodynamic cosmological simulations (Miralda-Escudé et al, 1996) and complimentary analytic treatments (e.g. Hui and Gnedin, 1997). The Ly α clouds were replaced by the Cosmic Web (Figure 4), the filamentary network of dark matter and baryons that describes the large-scale structure of a CDM universe. The HI Ly α forest traces the undulations in this web and this so-called fluctuating Gunn-Peterson approximation offers a terrific description of the IGM with sound analytic underpinnings.

Figure 5 compares an early generation model of the IGM from a hydrodynamic simulation against a portion of a Keck/HIRES spectrum. The agreement is remarkable and even the expert reader is challenged to identify which panel is real and which is simulated. The cosmic web paradigm is a true triumph of CDM cosmology and its development ushered in the opportunity to leverage IGM observations for research in cosmology.

For the last decade, observational advances have stemmed largely from the massive spectroscopic surveys of the Sloan Digital Sky Survey (SDSS). These have yielded terrific statistical descriptions of the IGM (Lee et al, 2015) across large areas of the sky for experiments like measurements of Baryonic Acoustic Oscillations (Busca et al, 2013). Large surveys of optically thick HI gas have also been com-

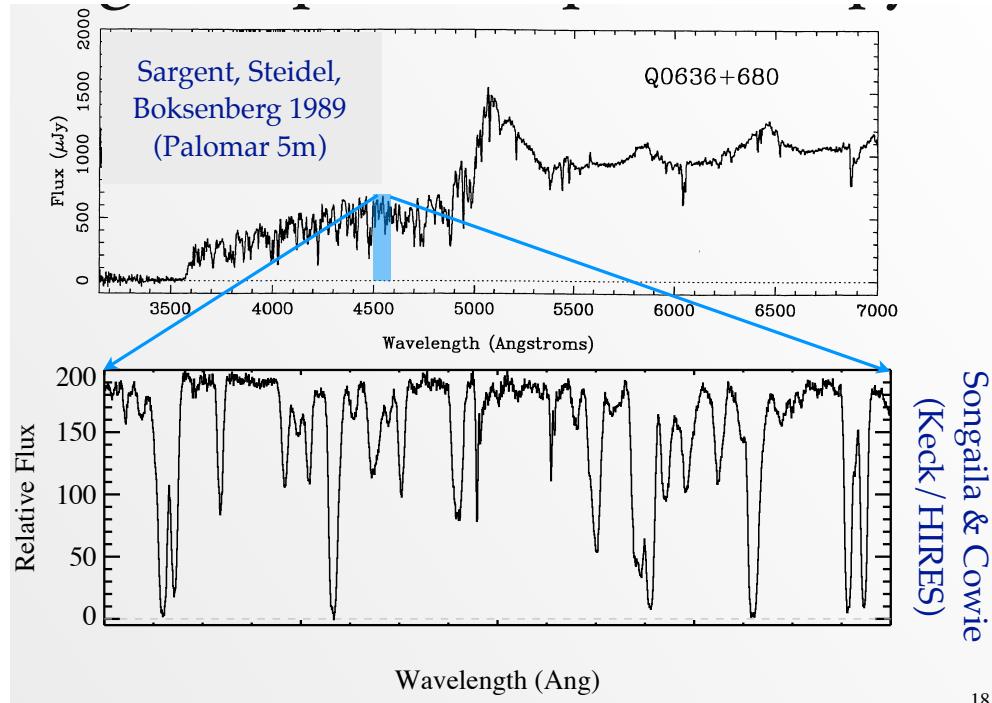


Fig. 3 Comparison of the high quality Palomar spectrum of Q0636+680 from Sargent et al (1989) against a snippet of spectra obtained with Keck/HIRES by Songaila and Cowie (2002). The latter fully resolves the HI Ly α absorption and, in essence, offers a complete description.

prised (Prochaska et al, 2005, 2010; Noterdaeme et al, 2012) and analysis probing the underlying dark matter density field probed by the IGM have emerged (Font-Ribera et al, 2012). In addition, the ongoing discovery of $z > 6$ quasars and GRBs coupled with high-performance echelle spectrometers have probed the IGM to the epoch of HI reionization. And, a series of increasingly sensitive UV spectrometers on the *Hubble Space Telescope* have anchored the results in the modern universe (Penton et al, 2000; Davé and Tripp, 2001).

This chapter is organized into the following sections: (i) the physics of HI Ly α absorption; (ii) key concepts of spectral-line analysis; (iii) characterizing the HI Ly α forest as absorption lines; (iv) optically thick HI absorption; and (v) a brief overview of modern analysis and results. The primary goal of these lectures was to provide the foundation and background for a young scientist to engage in empirical analysis of the IGM. The focus throughout is observational and the approaches are largely traditional; excellent reviews with a greater emphasis on theory are given by Meiksin (2009) and McQuinn (2016).

This Chapter is also supplemented by the lecture notes and slides presented in SaasFee, and a set of iPython Notebooks illustrating concepts and providing example code for related calculations and modeling. These supplementary materials are publicly available at <https://github.com/profxj/SaasFee2016>. Python code relevant to HI Ly α absorption and IGM analysis are packaged as `linetools`² and `pyigm`³ on github.

Before continuing, I offer a few caveats on this work: (1) I have not attempted to keep pace with the new literature (since the School); (2) I have only ingested a portion of the DLA lecture. See Wolfe et al (2005) for a fine review; (3) I have only summarized the slides that ended the series of lectures. The modern literature is evolving rapidly with numerous on-going and upcoming experiments (e.g. eBOSS, CLAMATO, DESI, PFS).

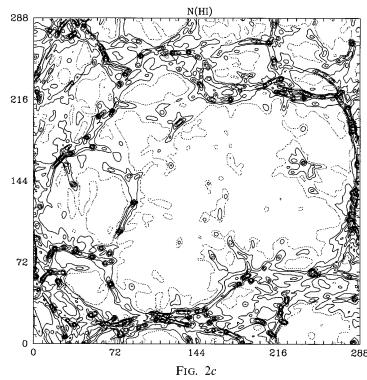


Fig. 4 Early illustration of the cosmic web from hydrodynamic simulations of the universe (Miralda-Escudé et al, 1996). These authors were among the first to associate the Ly α forest with the underlying, dark matter density field.

² <https://github.com/linetools/linetools>

³ <https://github.com/pyigm/pyigm>

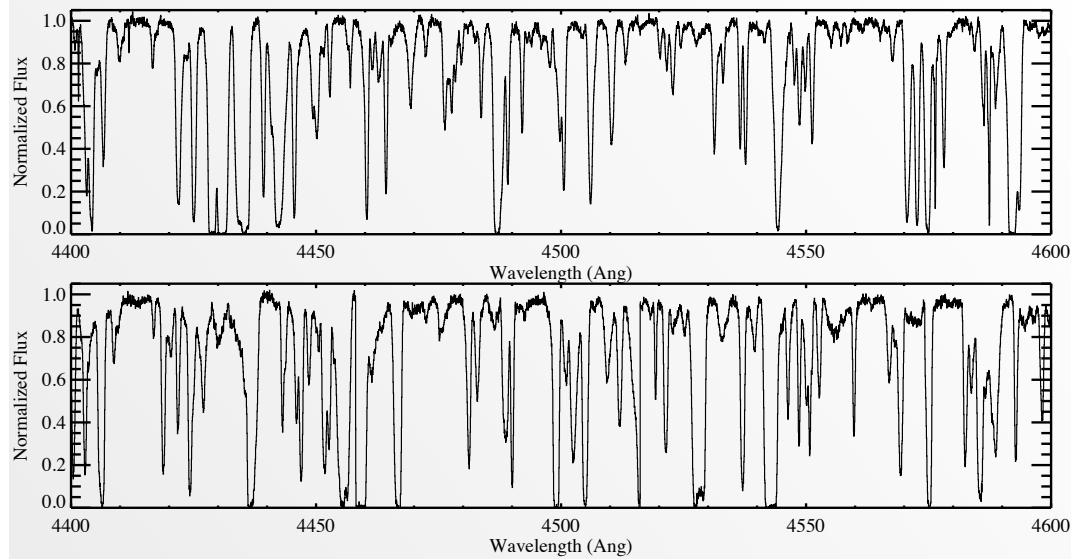


Fig. 5 Comparison of a snapshot of Ly α forest spectra (obtained with the Keck/HIRES spectrometer; (Kirkman and Tytler, 1997)) with a numerical simulation of the Ly α forest (Miralda-Escudé et al, 2000). Even an expert struggles to distinguish one from the other. Can you tell which is which?!

2 Physics of Lyman Series Absorption

The HI Ly α photon may be emitted following one of several processes: (i) the resonant absorption of a Ly α photon by atomic hydrogen; (ii) as the final emission in the recombination cascade of hydrogen; (iii) following the collisional excitation of HI. In this Chapter, we concern ourselves with the first process – HI Ly α resonant-line scattering – although the other processes may play a critical role in Ly α radiative transfer. In this section we will describe the physics of HI line absorption, introduce the concepts of a line-profile, and illustrate the basics of Ly α absorption lines. This section is supplemented by the iPython *HILyman_Series.ipynb* Notebook and the "sassfee16_lymanseries" Lecture. These are available online: <https://github.com/profxj/SaasFee2016>.

2.1 HI Energy Levels

We begin with a derivation of the energy levels for atomic hydrogen. Energies in the classic Rutherford-Bohr model of a Hydrogenic ion with charge Ze are solved from a standard Hamiltonian $H^{(0)}$ with an electrostatic potential

$$H^{(0)} = \frac{-\hbar\nabla^2}{2m} - \frac{Ze^2}{r} . \quad (2)$$

For energy level n , one recovers

$$E_n = -\frac{1}{2}\mu c^2 \frac{(Z\alpha)^2}{n^2} , \quad (3)$$

and the quantum states described by n, ℓ, m, m_s or $|n\ell mm_s\rangle$ are degenerate in ℓ, m, m_s because our Hamiltonian is rotationally invariant. In Equation 3, we have the fine-structure constant $\alpha \equiv e^2/\hbar c \approx 1/137$ and the reduced mass $\mu = \frac{m_e(Zm_p)}{m_e + Zm_p}$. For Hydrogen, $\mu \approx 0.999m_e$.

From E_n , we may evaluate the wavelengths for the Lyman series (transitions linked to the ground state E_1), as

$$\lambda_{rest,n} = \frac{hc}{E_n - E_1} . \quad (4)$$

Table 1 lists the calculated values from Equations 3 and 4 for ≈ 20 Lyman series lines, compared against empirical measurements. One identifies a systematic offset of $\delta\lambda \approx 0.015\text{\AA}$ between the Rutherford-Bohr energies and experiment. These result from perturbations to the standard Hamiltonian that we now consider.

Table 1 HI Lyman Series Lines

Transition	n	$E_n - E_1$ (eV)	λ_{rest} (\text{\AA})	λ_{exp} (\text{\AA})
Ly α	2	10.200000	1215.6845	1215.6701
Ly β	3	12.088889	1025.7338	1025.7223
Ly γ	4	12.750000	972.54759	972.5368
Ly δ	5	13.056000	949.75351	949.7431
Ly ϵ	6	13.222223	937.81375	937.8035
Ly ζ	7	13.322449	930.75844	930.7483
Ly η	8	13.387500	926.23580	926.2257
Ly θ	9	13.432099	923.16041	923.1504
Ly ι	10	13.464000	920.97310	920.9631
Ly κ	11	13.487604	919.36139	919.3514
Ly λ	12	13.505556	918.13933	918.1294
Ly μ	13	13.519527	917.19053	917.1806
Ly ν	14	13.530613	916.43908	916.429
Ly ρ	15	13.539556	915.83374	915.824
Ly σ	16	13.546875	915.33891	915.329
Ly τ	17	13.552942	914.92921	914.919
Ly χ	18	13.558025	914.58616	914.576
Ly ψ	19	13.562327	914.29604	914.286
Ly ω	20	13.566000	914.04849	914.039

There are two perturbations with energies that scale as α^4 , i.e. the next terms in Equation 3: (1) spin-orbit coupling and (2) the first expansion of the relativistic kinetic energy. Classically, the spin-orbit coupling is described as a magnetic dipole

interaction between the spin of the electron and the orbit of the nucleus, i.e. the e^- observes a magnetic field due to the current driven by the nucleus:

$$\mathbf{B} = -\frac{1}{c} \mathbf{v} \times \mathbf{E} = \frac{1}{m_e c r} \ell \frac{d\phi}{dr} , \quad (5)$$

with energy

$$E = -\mu_s \cdot \mathbf{B} \quad \text{with} \quad \mu_s = -\frac{eg\mathbf{s}}{2m_e c} , \quad (6)$$

where $g \approx 2$ for an electron. The Hamiltonian for the perturbation is:

$$H_{SO} = \frac{1}{2m_e^2 c^2} \mathbf{S} \cdot \mathbf{L} \frac{1}{r} \frac{d\phi}{dr} , \quad (7)$$

and the standard treatment is to identify an operator that commutes with $H^{(0)}$ and H_{SO} , and also uniquely identifies the degenerate states. We choose $\mathbf{J} \equiv \mathbf{L} + \mathbf{S}$ recognizing

$$\mathbf{L} \cdot \mathbf{S} = \frac{1}{2} (|\mathbf{J}|^2 - |\mathbf{L}|^2 - |\mathbf{S}|^2)$$

which yields energies

$$E_{SO} = \langle H_{SO} \rangle = \frac{1}{2} C [j(j+1) - \ell(\ell+1) - s(s+1)] \quad (8)$$

with C a constant. For fixed \mathbf{L} and \mathbf{S} (i.e. splitting within a level, e.g. 2P), we have $\Delta E_{SO} = E_{J+1} - E_J = C(j+1)$. For our Hydrogenic ion with $\phi = Ze^2/r$,

$$\frac{d\phi}{dr} = \frac{-Ze^2}{r^2} \quad (9)$$

and

$$H_{SO} = \frac{Ze^2}{2m_e^2 c^2} \frac{1}{r^3} \mathbf{L} \cdot \mathbf{S} . \quad (10)$$

Following standard perturbation theory⁴, we find

$$\langle H_{SO} \rangle = -E_n \frac{Z^2 \alpha^2}{2n} \frac{[j(j+1) - \ell(\ell+1) - s(s+1)]}{\ell(\ell+\frac{1}{2})(\ell+1)} \quad (11)$$

with E_n given by Equation 3. As advertised, spin-orbit coupling is 4th order in α and we now have explicit energy dependence on j, ℓ and s . For the Hydrogen $n=2$ levels ($Z=1; n=2; \ell=0, 1; j=1/2, 3/2$), we find

⁴ See the Lecture notes for an expanded derivation.

$$\langle H_{SO} \rangle = \begin{cases} 0 & 2^2S_{\frac{1}{2}} \quad (\ell = 0; j = 0) \\ \frac{mc^2\alpha^4}{96} & 2^2P_{3/2} \quad (\ell = 1; j = 3/2) \\ -\frac{mc^2\alpha^4}{48} & 2^2P_{1/2} \quad (\ell = 1; j = 1/2) \end{cases} \quad (12)$$

giving a $2P_{3/2}$ - $2P_{1/2}$ splitting of 4.5×10^{-5} eV or $\Delta v \approx \Delta E/cE \approx 1$ km/s.

To derive the Relativistic correction to order α^4 , we expand the K.E. to the next term in v^2/c^2 from the Lagrangian

$$\text{K.E.} = \frac{p^2}{2m} \left(1 - \frac{1}{4} \frac{v^2}{c^2} \right). \quad (13)$$

This gives a relativistic perturbation

$$H_{rel} = -\frac{1}{2mc^2} \left(\frac{p^2}{2m} \right)^2 \quad (14)$$

that has no spin dependence (spherically symmetric) such that $[H_{rel}, L^2] = [H_{rel}, L] = 0$ and the standard $|n\ell mm_s\rangle$ diagonalize H_{rel} . If we recognize that

$$H_{rel} = -\frac{1}{2mc^2} \left(H^{(0)} - V^{(0)} \right)^2$$

with $V^{(0)} = -Ze^2/r$, it is straightforward to compute the energies

$$\langle H_{rel} \rangle = -E_n \frac{Z^2\alpha^2}{n} \left(\frac{3}{4n} - \frac{1}{\ell + \frac{1}{2}} \right) \quad (15)$$

Combining this result with spin-orbit coupling, we recover a remarkable expression that depends only on j :

$$\langle H_{SO} \rangle + \langle H_{rel} \rangle = E_n \frac{Z^2\alpha^2}{n} \left[\frac{1}{j + \frac{1}{2}} - \frac{3}{4n} \right] \quad (16)$$

Altogether, the α^4 term for E_n has no explicit ℓ dependence, nor any s dependence. Furthermore, higher j implies higher energy following the 3rd Hund's rule. The energy shifts from E_n for Hydrogen are then

$$\Delta E_{nj} = -7.25 \times 10^{-4} \text{eV} \frac{1}{n^3} \left[\frac{1}{j + \frac{1}{2}} - \frac{3}{4n} \right] \quad (17)$$

which for the $n = 1, 2$ states of Hydrogen evaluate to the results in Table 2.

The splitting of the $n = 2$ level implies that Ly α (1S-2P) is a doublet with ≈ 1 km/s separation which is generally too small to resolve observationally but could be important for radiative transfer treatments.

Returning to Table 1, the implied shift from the Rutherford-Bohr energies $\delta\lambda/\lambda \sim \delta E/E$ is calculated using δE from Table 2 and accounting for the relative degeneracy of the $2^2P_{\frac{1}{2}}, 2^2P_{\frac{3}{2}}$ states yields

Table 2 Perturbations to the $n = 1, 2$ levels of Hydrogen

State	n	j	$\langle H_{SO} \rangle$	$\langle H_{rel} \rangle$
$1^2S_{\frac{1}{2}}$	1	$\frac{1}{2}$	-1.8×10^{-4} eV	
$2^2S_{\frac{1}{2}}, 2^2P_{\frac{1}{2}}$	2	$\frac{1}{2}$	-5.7×10^{-5} eV	
$2^2P_{\frac{3}{2}}$	2	$\frac{3}{2}$	-1.1×10^{-5} eV	

$$\delta E = \frac{1}{3} \left[2\delta E_{1S \rightarrow 2P_{\frac{3}{2}}} + \delta E_{1S \rightarrow 2P_{\frac{1}{2}}} \right] \quad (18)$$

$$= 1.55 \times 10^{-4} \text{ eV} \quad (19)$$

or

$$\Delta \lambda = -\frac{\delta E}{E} \times 1215.68 \text{ Å} = -0.0138 \text{ Å} . \quad (20)$$

Voila!

2.2 The Line Profile

We now derive the line profile of HI Lyman series absorption which expresses the energy dependence of the photon cross-section. This results from two physical effects: (1) the quantum mechanical coupling of the energy levels described in the previous sub-section; (2) Doppler broadening from the kinetic motions of the gas. We reserve a discussion of the observed line-profile related to the instrument to the next section.

We express the opacity $\kappa_{jk}(\nu)$ of a gas with number density n_j in state j as:

$$\kappa_{jk}(\nu) = n_j \sigma_{jk}(\nu) \quad (21)$$

with σ_ν the photon cross-section at frequency ν for a transition to state k . We separate the frequency dependence by introducing the line-profile $\phi(\nu)$

$$\sigma(\nu) = \sigma_{jk} \phi(\nu) \quad (22)$$

with σ_{jk} the integrated cross-section over all frequencies and $\phi_\nu d\nu$ reflects the probability an atom will absorb a photon with energy in $\nu, \nu + d\nu$.

A naive guess for $\phi(\nu)$ is the Delta function, i.e. the transition occurs only at the exact energy splitting the energy levels from the previous subsection:

$$\phi_\delta(\nu) = \delta(\nu - \nu_{jk}) \quad (23)$$

Quantum mechanically, however, the excited $n = 2$ state has a half-life $\tau_{\frac{1}{2}}$ given by the Spontaneous emission coefficient, $\tau_{\frac{1}{2}} = 1/A_{jk}$ ($A_{jk} = 6.265 \times 10^8 \text{ s}^{-1}$ for HI Ly α). This finite lifetime implies a finite width ΔE to the energy level which may be estimated from the Heisenberg Uncertainty principle

$$\Delta E \sim \frac{\hbar}{\Delta t} \sim \hbar A . \quad (24)$$

This demands a finite width to the line-profile. Define $W_{jk}(E)$ as the quantum mechanical probability of a transition occurring between states j and k with energy E , and $W_j(E)$ as the probability of state j being characterized by the energy interval $(E_j, E_j + dE_j)$. A standard Quantum mechanical treatment shows that $W_j(E)$ has a Lorentzian shape (aka the Breit-Wigner profile):

$$W_j(E_j)dE_j = \frac{\gamma_j dE_j/h}{(2\pi/h)^2 [E_j - < E_j >]^2 + (\gamma_j/2)^2} \quad (25)$$

with

$$\gamma_j \equiv \sum_{i < j} A_{ij} \quad (26)$$

For a coupling between only two states j, k , one derives W_{jk} by convolving W_j and W_k :

$$W_{jk}(E)dE = \frac{[\gamma_j + \gamma_k] dE/h}{(2\pi/h)^2 [E - E_{jk}]^2 + ([\gamma_j + \gamma_k]/2)^2} \quad (27)$$

We reemphasize that this calculation was restricted to the j and k states whereas a proper calculation needs to consider the coupling of all the energy levels. For our Lyman series lines we note that $\gamma_j = 0$ as, by definition, there are no energy levels below the ground state.

From Equation 27, we introduce the Natural line-profile normalized to have unit integral value in frequency,

$$\phi_N(\nu) = \frac{1}{\pi} \left[\frac{(\gamma_j + \gamma_k)/4\pi}{(\nu - \nu_{jk})^2 + (\gamma_j + \gamma_k)^2/(4\pi)^2} \right] . \quad (28)$$

giving (at last)

$$\sigma(\nu) = \sigma_{jk}\phi_N(\nu) \quad (29)$$

As illustrated in the Notebook, the "wings" of $\phi_N(\nu)$ are ≈ 10 orders of magnitude down from line-center. Remarkably these can be very important for Ly α .

It is standard practice to express the normalization σ_{jk} in terms of the oscillator strength f_{jk} which is either measured empirically (preferred) or estimated theoretically

$$\sigma_{jk} = \frac{\pi e^2}{m_e c} f_{jk} . \quad (30)$$

Our final expression becomes

$$\sigma_v = \frac{\pi e^2}{m_e c} f_{jk} \left[\frac{(\gamma_j + \gamma_k)/4\pi^2}{(v - v_{jk})^2 + (\gamma_j + \gamma_k)^2/(4\pi)^2} \right] \quad (31)$$

Expressing the FWHM of the line-profile as the frequency width where $\sigma(v)/\sigma(v)_{max} = 1/2$, we have $\Delta v_{FWHM} = \pm \frac{\gamma_j + \gamma_k}{4\pi}$ which very nearly matches our estimate from the Uncertainty Principle! For HI Ly α with $\gamma_1 = 0, \gamma_2 = A_{21}$, we find the FWHM in velocity to be

$$\Delta v_{FWHM} = c \Delta v_{FWHM} / v \approx 1.5 \times 10^{-2} \text{ km/s} \quad (32)$$

For astrophysical purposes, this nearly is a delta function.

In an astrophysical environment, each atom in a gas has its own motion which spreads the line without changing the total amount of absorption. This Doppler effect, to lowest order in v/c , is

$$\Delta v = v - v_{jk} = v_{jk} \frac{v}{c} \quad (33)$$

Assuming first that the gas motions are characterized solely by T (i.e. no turbulence), we adopt a Maxwellian distribution for particles of mass m_A giving a profile function:

$$\phi_D(v) = \frac{1}{\Delta v_D \sqrt{\pi}} \exp \left[-\frac{(v - v_{jk})^2}{\Delta v_D^2} \right] \quad (34)$$

$$\text{with } \Delta v_D \equiv \frac{v_{jk}}{c} \sqrt{\frac{2kT}{mA}} \quad (35)$$

At line-center ($v = v_{jk}$), the cross-section (neglecting stimulated emission) is

$$(\sigma_v^D)_{max} = \sigma_{jk} \phi_D(v_{jk}) = \frac{\pi e^2}{mc} f_{jk} \frac{1}{\Delta v_D \sqrt{\pi}} \quad (36)$$

This cross-section is several orders of magnitude lower than $(\sigma_v^N)_{max}$ as the absorption has been ‘spread’ over a velocity interval several orders of magnitude larger than the width estimated by Equation 32.

We can generalize the profile to include random, turbulent motions (characterized by ξ) by modifying the Doppler width,

$$\Delta v_D = \frac{v_{jk}}{c} \left(\frac{2kT}{mA} + \xi^2 \right)^{\frac{1}{2}} . \quad (37)$$

Expressing the line-profile in velocity space, we introduce the Doppler parameter

$$b \equiv \sqrt{\frac{2kT}{m_A} + \xi^2} \quad (38)$$

and the velocity line-profile for Doppler motions is

$$\phi_D(v) = \frac{1}{b\sqrt{\pi}} \exp\left[-\frac{v^2}{b^2}\right] . \quad (39)$$

See the Notebook for a series of examples illustrating this profile in comparison to the Natural profile.

Generally, the cross-section has contributions from both Natural and Doppler broadening, with Doppler broadening dominating the line-center and the Lorentzian of Natural broadening dominates the wings. The overall profile is a convolution of the two terms,

$$\phi_V(v) = \frac{\gamma}{4\pi} \int_{-\infty}^{\infty} \frac{\left(\frac{m}{2\pi kT}\right)^{\frac{1}{2}} \exp\left(-\frac{mv^2}{2kT}\right)}{(v - v_{jk} - v_{jk}v/c)^2 + (\gamma/4\pi)^2} dv \quad (40)$$

and one is inspired to introduce the Voigt function

$$H(a, u) = \frac{a}{\pi} \int_{-\infty}^{\infty} \frac{e^{-y^2}}{a^2 + (u - y)^2} dy \quad (41)$$

and we identify a, u as:

$$a \equiv \frac{\gamma}{4\pi \Delta v_D} \quad (42)$$

$$u \equiv \frac{v - v_{jk}}{\Delta v_D} \quad (43)$$

Altogether, we have

$$\phi_V(v) = \frac{H(a, u)}{\Delta v_D \sqrt{\pi}} \quad (44)$$

which has no analytic solution. For speed, one often relies on look-up tables. In Python, the real part of `scipy.special.wofz` is both accurate and fast (see the Voigt documentation in the `linetools` package).

The Lorentzian profile from Natural broadening is only an approximation because it ignores the coupling between states other than j, k . Scattering is a second-order quantum mechanical process: annihilation of one photon and the creation of a scattered photon. A proper treatment requires second-order time dependent perturbation theory and requires one to sum over all bound-states and integrate over all continuum-state contributions. Lee (2003) have calculated a series expansion of the corrections to the Voigt profile and we refer the reader to his paper for a more detailed presentation. Here, we report the next term in the series for HI Ly α :

$$\sigma_v = \sigma_T \left(\frac{f_{jk}}{2} \right)^2 \left(\frac{v_{jk}}{\delta v} \right)^2 \left[1 - 1.792 \frac{\delta v}{v_{jk}} \right] . \quad (45)$$

with σ_T the Thompson cross-section. It is evident that the correction is not symmetric about line-center. This leads to an asymmetry which shifts the measured line-center if the gas opacity is very high (e.g. $N_{\text{HI}} > 10^{21.7} \text{ cm}^{-2}$).

2.3 Optical Depth (τ_v) and Column Density (N)

We now introduce two quantities central to absorption-line analysis. First, the optical depth τ_v which is defined as the integrated opacity along a sightline. In differential form $d\tau(v) = -\kappa(v)ds$ implying

$$\tau(v) = \sigma(v) \int n_j ds \quad (46)$$

which gives an explicit frequency dependence related to the line-profiles of the previous sub-section.

We recognize the second term to be the column density N_j ,

$$N_j \equiv \int n_j ds \quad (47)$$

which has units of cm^{-2} and is akin to a surface density ($n_j \rightarrow \rho$; $N_j \rightarrow \Sigma$). As an example, consider the column density of O₂ through 1 m of air. With $\rho_{\text{O}_2} = 1.492 \text{ g/L}$, $N_{\text{O}_2} = 3 \times 10^{21} \text{ cm}^{-2}$.

A quantity of particular interest is the optical depth at line-center τ_0 of a gas with N_j and Doppler parameter b . At line-center ($v = v_{jk}$), our line-profile is dominated by Doppler motions $\phi_v(v_{jk}) \approx \phi_D(v_{jk})$ and from Equation 39, we recover

$$\tau_0 = \frac{\sqrt{\pi} e^2}{m_e c} \frac{N_j \lambda_{jk} f_{jk}}{b} . \quad (48)$$

For Ly α , with b expressed in km/s:

$$\tau_0^{\text{Ly}\alpha} = 7.6 \times 10^{-13} \frac{N[\text{cm}^{-2}]}{b[\text{km/s}]} \quad (49)$$

2.4 Idealized Absorption Lines

Without derivation (see the Lecture notes), we express the radiative transfer for a source with intensity I_v^* through a medium with optical depth τ_v as

$$I_v = I_v^* e^{-\tau_v} \quad (50)$$

Therefore, we may consider the formation of absorption lines as the simple integration of the optical depth.

In Figure 6, we show a series of idealized HI Ly α lines for a range of HI column densities and Doppler parameters (N_{HI} , b). These illustrate the shapes of the line-profiles, here dominated by Doppler broadening, and the varying optical depth at line-center. These are plotted in velocity space, taking $\delta v = c\delta\lambda/\lambda_0$ with $\delta v = 0$ km/s corresponding to the line-center. In Figure 7 we contrast lines which have $\tau_0 \approx 1$, with an HI absorption line with very high N_{HI} and correspondingly high τ_0 . Here the line profile is entirely determined by Natural broadening.

2.5 Equivalent Width

The equivalent width for absorption W_λ is a gross measure of the flux absorbed (scattered) by the gas cloud. Strictly, W_λ is the convolution of the optical depth with the line profile. And although it is primarily an observational quantity, its value is *independent* of the instrument profile and it does depend on physical properties of the absorbing gas.

Analytically, we define

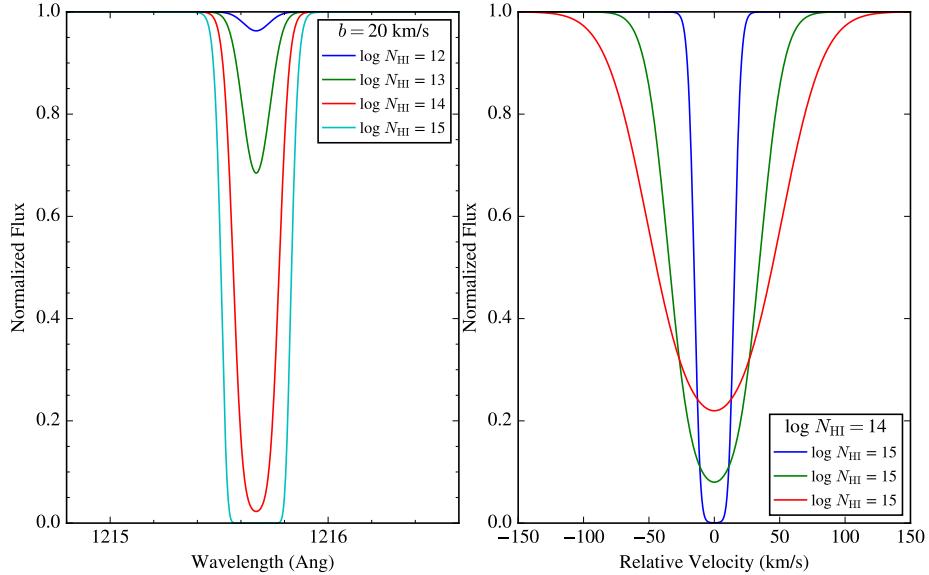


Fig. 6 A series of absorption-line profiles for HI Ly α for a gas with varying column density. These are presented as a function of wavelength (left) and velocity (right) with the latter given by $\delta v = c\delta\lambda/\lambda_0$.

$$W_\lambda = \int_0^\infty \left[1 - \frac{I_\nu}{I_\nu^*} \right] d\lambda \quad (51)$$

Substituting our simple radiative transfer equation, this gives

$$W_\lambda = \int_0^\infty [1 - \exp(-\tau_\lambda)] d\lambda \quad (52)$$

One may contrast this equation with analysis of stellar atmospheres which has a very different radiative transfer equation. The W_λ value, expressed in Å, may be visualized as the width of a box-car profile that matches the absorbed flux in a normalized spectrum, e.g. Figure 8.

2.6 Curve of Growth for HI Ly α

It is valuable to develop an intuition of the relation between equivalent width and the physical properties of a gas (N, b). This relationship is generally referred to as the curve-of-growth (COG) and may be inverted to constrain N, b from W_λ measurements. The COG also nicely describes the transition from a Doppler-dominated line to a naturally-broadened line.

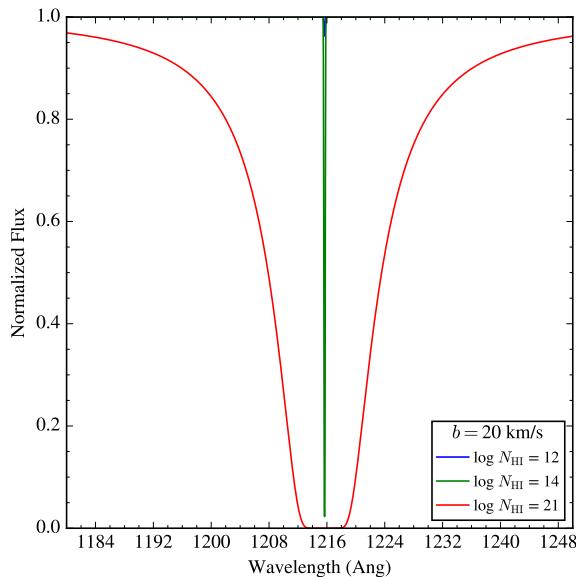


Fig. 7 Comparison of absorption for gas with optical depth near unit (blue/green lines) to gas exhibiting damped Ly α absorption (red).

One typically considers three regimes for the COG which depend on the central optical depth of the absorption line τ_0 (Equation 48). In the Weak limit ($\tau_0 \ll 1$), Natural broadening is negligible and Doppler broadening dominates,

$$\tau_v = \frac{\pi e^2}{m_e c} \lambda f_{jk} N_j \phi_D(v) . \quad (53)$$

With τ_0 small, Equation 52 reduces to

$$W_\lambda = \frac{\lambda^2}{c} \int [1 - \exp(-\tau_v)] dv \quad (54)$$

$$\approx \frac{\lambda^2}{c} \int_0^\infty \tau_v dv \quad (55)$$

$$= \frac{\lambda^2}{c} \frac{\pi e^2}{m_e c} f_{jk} N_j \quad (56)$$

revealing a linear relationship between W_λ and N_j and no dependence on b . The Weak limit, therefore, is also referred to as the ‘linear’ portion of the COG. Evaluating for Ly α , we find

$$W_\lambda^{\text{Ly}\alpha} \approx (0.1 \text{ \AA}) \frac{N_{\text{HI}}}{1.83 \times 10^{13} \text{ cm}^{-2}} \quad (57)$$

Our estimate of τ_0 (Equation 48) requires $N_{\text{HI}} \ll 10^{14} \text{ cm}^{-2}$ for $\tau_0 \ll 1$, which also implies $W_\lambda \ll 1 \text{ \AA}$. This column density is many orders of magnitude less than that typical of the Galactic interstellar medium (ISM). It implies a gas that is extremely diffuse and, likely, highly ionized. These are the physical conditions of the IGM.

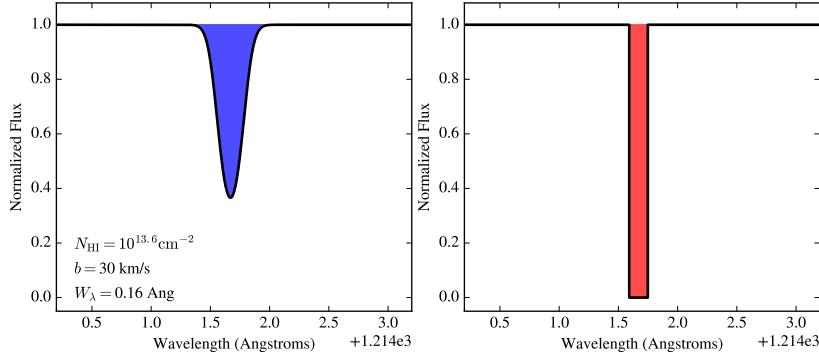


Fig. 8 (left) An HI Ly α absorption line with $N_{\text{HI}} = 10^{13.6} \text{ cm}^{-2}$ and $b = 30 \text{ km/s}$. Integrating the area above the curve (to the normalized continuum), one recovers an equivalent width $W_\lambda = 0.16 \text{ \AA}$. (right) A cartoon representation of the corresponding boxcar profile giving the same equivalent width.

The Strong line limit of the COG refers to $\tau_0 \gtrsim 1$ and we may still ignore Natural broadening. In this regime, the incident flux is strongly absorbed at line-center and the equivalent width is well described by the width of the line (i.e. it is nearly approximated as a ‘box’; Figure 9). Expressing the frequency dependence of the optical depth as $\tau_x = \tau_0 e^{-x^2}$, with $x \equiv \Delta v / \Delta v_D$, we may estimate the line width by considering the value of x that gives $\tau_x = 1$. Trivially, we find $x_1 = \sqrt{\ln \tau_0}$ and therefore

$$W_\lambda \approx 2x_1 \approx 2\sqrt{\ln \tau_0} \quad (58)$$

in the Strong regime, also known as the saturated limit. To increase W_λ in this saturated limit, we need to increase τ_0 immensely and likewise N , i.e. $W_\lambda \propto (\ln N)^{\frac{1}{2}}$. Whereas W_λ is insensitive to τ_0 , it is sensitive to the internal structure of the cloud $W_\lambda \propto x \propto b$.

Lastly, there is the Damping regime with $\tau_0 \gg 1$ and where Natural broadening dominates. Now the optical depth is given by the Lorentzian profile

$$\tau_x \approx \frac{\tau_0 A}{\sqrt{\pi}} \frac{1}{x^2} \quad (59)$$

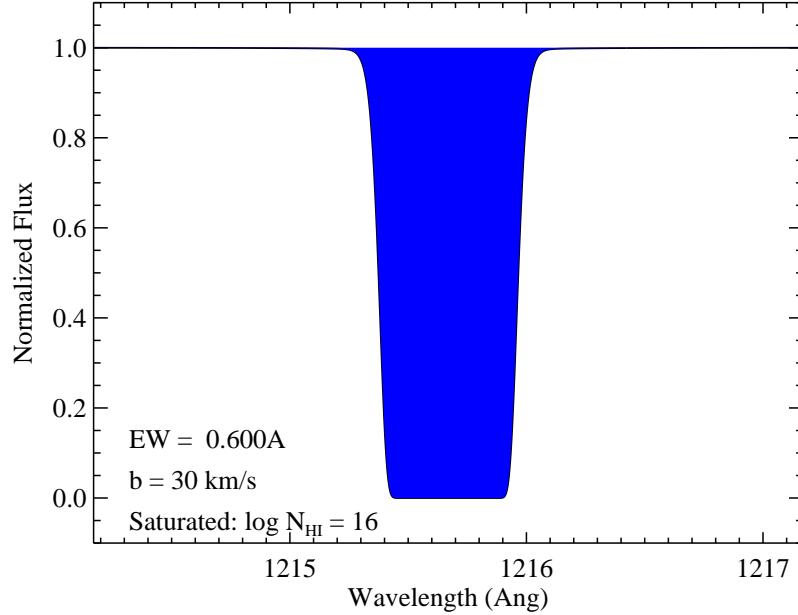


Fig. 9 Saturated Ly α line in the Strong regime of the COG. Here the optical depth at line-center τ_0 exceeds unity such that the incident photons are completely absorbed. But, the opacity is insufficient for the Lorentzian wings to appear; the line-profile is instead dominated by the Maxwellian of Doppler broadening (i.e. Gaussian) giving a box-like appearance.

and the width (estimated from $\tau_x = 1$) is $x_1 = [\tau_0 A]^{\frac{1}{2}}$. Therefore, $W_\lambda \approx 2x_1 \propto N^{\frac{1}{2}}$ or formally

$$\frac{W_\lambda}{\lambda} \approx \frac{2}{c} \left[\lambda^2 N_j \frac{\pi e^2}{m_e c} f_{jkA} \right]^{\frac{1}{2}} \quad (60)$$

In the Damping regime, the equivalent width scales as \sqrt{N} with no dependence on the Doppler parameter as the core is fully saturated.

Figure 10 shows the COG for a single HI Ly α line with varying N_{HI} and Doppler parameter. The three COG regimes are well-described. We stress that only a small portion of the Weak limit ($W_\lambda \ll 1\text{\AA}$) permits actual detections with modern spectrometers and that the Damping regime is limited to gas with galactic surface densities. This leaves approximately 6 orders of magnitude in N_{HI} in the Strong (saturated) regime where observations of the equivalent width for Ly α alone offer a weak constrain on the gas column density.

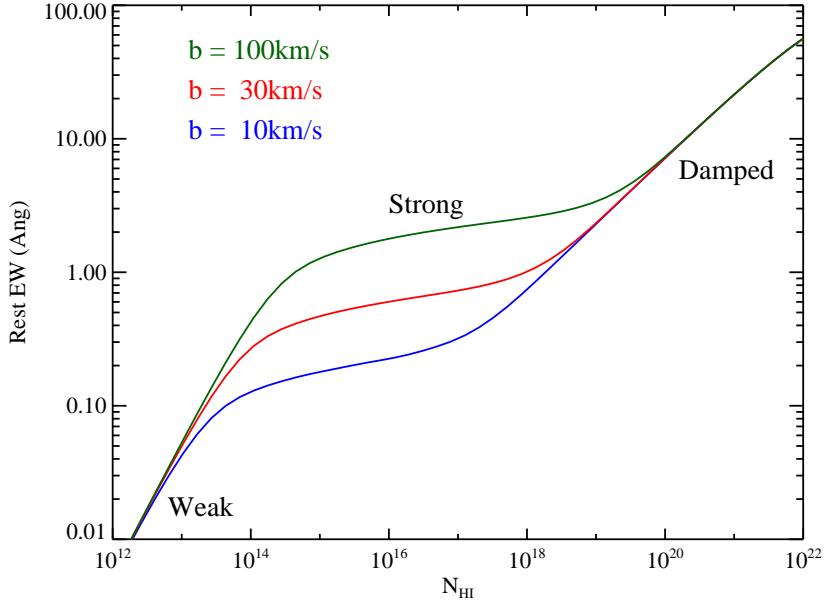


Fig. 10 Curve of growth (COG) for the HI Ly α line. Expressed across 10 orders of magnitude, one traverses from the weak ($\tau_0 \ll 1$), to the strong ($\tau_0 \sim 1$), to the damped regimes ($\tau_0 \gg 1$). Note the strong dependence on the Doppler parameter only in the strong regime.

2.7 Curve of Growth for the Lyman Series

Another application of the COG is to evaluate the absorption from a series of transitions from a single ion in a single ‘cloud’ of gas. Such analysis may enable one to derive more precisely the physical parameters (N, b) of the absorbing gas. Recalling that

$$\tau_0 = \frac{\sqrt{\pi}e^2}{m_e c} \frac{\lambda_{jk} f_{jk} N_j}{b} , \quad (61)$$

a single hydrogen gas cloud with fixed N, b will show decreasing τ_0 in increasing terms of the Lyman series. Therefore, the Lyman series absorption generates a COG that may span several of the regimes described above.

The standard analysis is to fit COG curves to a series of W_λ measurements to constrain N, b as illustrated in Figure 11. Another example and related code are provided in the Notebook.

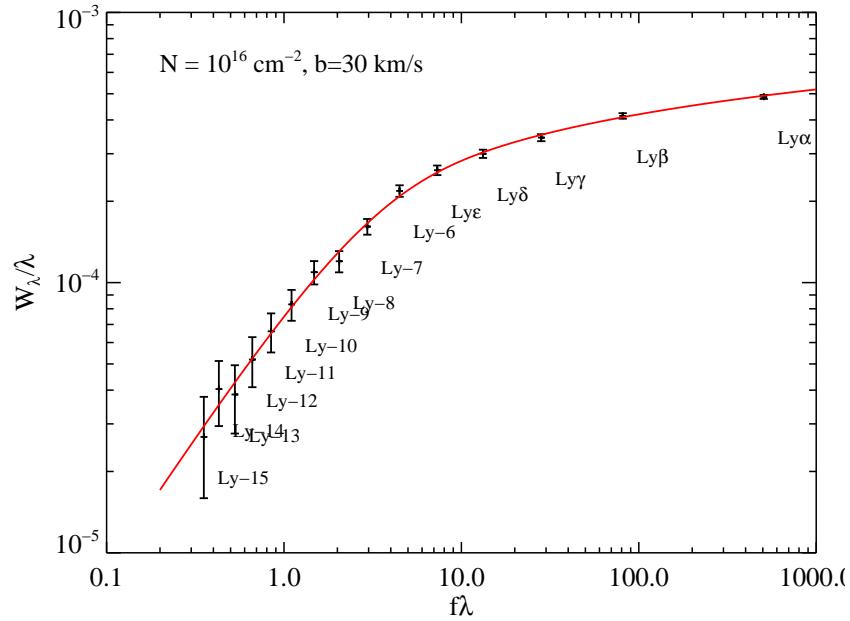


Fig. 11 The points are a set of equivalent width measurements for an HI gas for a series of HI Lyman lines. These are plotted against $f\lambda$ which is proportional to τ_0 . The red curve is a COG fit to the data, parameterized by N and b of the gas.

3 Basics of Spectral Analysis (for Absorption)

In this section, we provide a brief introduction to several key concepts of spectroscopy and then describe several of the fundamental aspects related to absorption-line analysis with data.

3.1 *Characteristics of a Spectrum*

It is beyond the scope of this Chapter to describe in detail the fundamentals of spectrometers in astronomy. For what follows, the reader need only appreciate that spectrometers' implement a dispersing element (e.g. grating) to redirect incident light as a function of wavelength such that a packet of photons with a distribution of energies are spread (monotonically) across a detector. The number of photons collected at a given detector pixel depends on the flux and the dispersion of the spectrograph. Of course, one cannot achieve infinite resolution and, further, the optics of the instrument affect the resultant spectral image.

One defines the Line Spread Function (LSF) as the functional form a monochromatic source exhibits on the detector. The diffraction of light (e.g. slit + dispersing element) transforms the input PSF of the source (typically a Moffat) into a line profile in wavelength. The precise LSF that results depends on type of spectrometer (e.g. Robertson, 2013) as summarized in Table 3.

Table 3 LINE SPREAD FUNCTIONS

Type of Spectrograph	LSF	R
Diffraction-limited slit	sinc^2	mN
Single-mode fiber	Gaussian	FWHM
Multi-mode fiber	Half ellipse	??
Fabry-Perot etalon	Lorentzian	mF

In Figure 12, we compare the idealized LSFs for the most common spectrometers in use: a diffraction-limited slit spectrometer and a fiber-fed spectrograph. Although the functional forms are distinct and one recognizes quantitative differences in the wings of the profiles, the inner line-profiles are very similar and one would require high signal data to distinguish the two. Furthermore, system imperfections in the spectrograph, variations in the PSF, and the Central Limit Theorem tends the LSF towards a Gaussian. For this school, we will adopt a Gaussian LSF. In practice, the LSF may be assessed empirically, e.g. through analysis of arc-line emission lamps and/or atmospheric absorption lines imprinted on an external source.

The primary impact of the LSF is to artificially broaden any spectral features through the convolution of the instrument profile with the intrinsic spectrum. While the LSF offers a complete description of this effect, we tend to describe it (and the

spectrometer) in terms of the width of the LSF, i.e. the resolution R . The actual metric for R depends on the shape of the LSF (i.e. Table 3) and for a Gaussian profile the standard metric is the full-width at half-maximum (FWHM). For an optical slit/fiber spectrograph, R is primarily dependent on the slit/fiber width (inversely proportional) and the grating. It is generally independent of any detector properties. Formally, we define

$$R \equiv \frac{\lambda}{\Delta\lambda_{\text{FWHM}}} \quad (62)$$

with λ the wavelength where the LSF has a measured FWHM of λ_{FWHM} .

While R describes the width of the LSF, the spectral width of a pixel on the detector is given by the dispersion $\delta\lambda$. Obviously, this depends on both R and the characteristics of the detector (i.e. the physical size of the pixel). To further confuse matters, many detectors can be ‘binned’ in electronics to increase $\delta\lambda$ and reduce the detector noise per effective pixel.

The sampling of the LSF is the number of pixels per resolution element, i.e. $\Delta\lambda_{\text{FWHM}}/\delta\lambda$. To achieve the full resolution of the spectrograph in the theoretical limit, one must sample the LSF by 2 or more pixels. This is commonly referred to as the Nyquist limit.

Commonly used spectrometers for HI absorption studies have R ranging from 2,000 (Keck/LRIS; SDSS) to many thousands (echellettes; Keck/ESI, VLT/X-Shooter), to tens of thousands (echelles; Keck/HIRES, VLT/UVES). Figure 13 compares spectra of the quasar FJ0812+32 observed with spectrometers having a range of R . It is evident that for the lower resolution spectra (SDSS), the intrinsic width of the lines have been substantially smeared by the instrumental LSF.

Although Equation 62 is the formal definition of R , astronomers frequently refer to only the FWHM of the LSF. In velocity, $\Delta v_{\text{FWHM}} = c/R$, e.g. $\Delta v_{\text{FWHM}} \approx 10 \text{ km/s}$

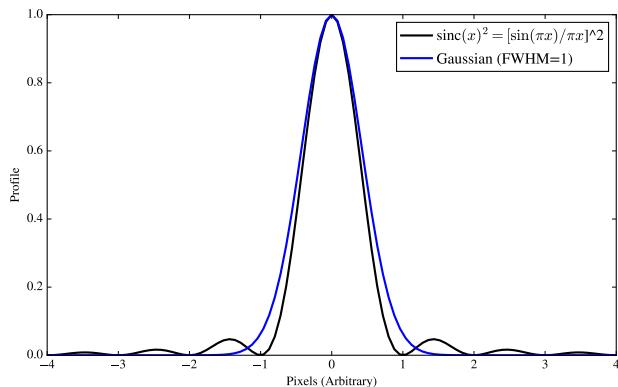


Fig. 12 Comparison of the LSF for a slit spectrometer (black) with that for a fiber-fed spectrometer (blue). Despite quantitative differences in the wings of the profiles, the majority of light is similarly described by the inner portion of the profiles.

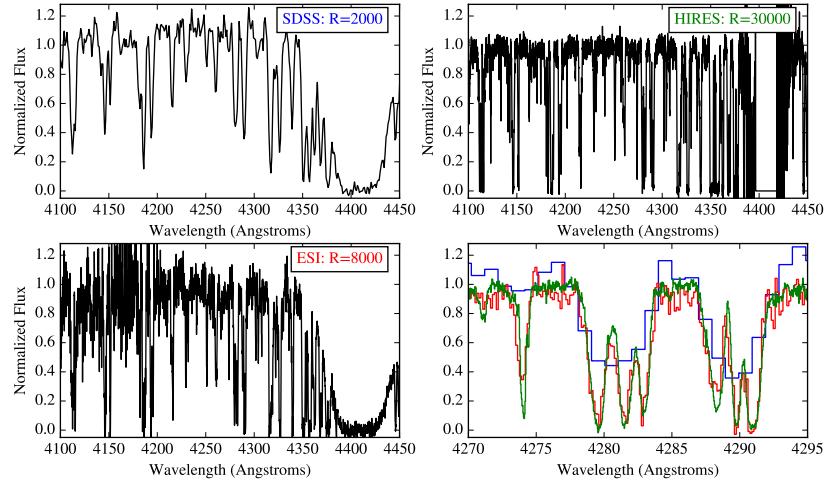


Fig. 13 Three spectra of the quasar J0812+32 observed with the SDSS, Keck/ESI, and Keck/HIRES spectrometers with resolution ranging from $R \approx 2,000 - 30,000$. The lower-right panel shows a zoom-in where one can easily appreciate the effects of varying spectral resolution.

for $R = 30,000$. This provides a physical intuition for the best sensitivity one achieves with the spectrometer. Old-timers (and many instrument web-pages) frequently refer to $\Delta\lambda_{\text{FWHM}}$, e.g. $\approx 2\text{\AA}$ for SDSS.

3.2 Continuum Normalization

Studies of HI in absorption require a background, light source with emission at FUV wavelengths (and bluer!) to excite HI at Ly α and higher order transitions. For extragalactic research, this has included the O and B stars in star-forming galaxies (Lee et al, 2014), the bright afterglows of gamma-ray bursts (Fynbo et al, 2009), and quasars (the focus of this Chapter). To estimate the opacity, one must estimate the source continuum f_λ^C then normalize the flux,

$$\bar{f}_\lambda = \frac{f_\lambda^{\text{obs}}}{f_\lambda^C} \quad (63)$$

In essence, we aim to remove any trace of the background source as it is (usually) scientifically irrelevant to the analysis. This process is referred to as continuum normalization and it can be the dominant systematic uncertainty in absorption analysis.

To date, quasars have been the most commonly observed background sources because they are the most common, steadily luminous, distant phenomena. Unfortunately, they may also exhibit the most complexity in their spectral energy distribution (SED). This includes very strong and wide emission lines, an underlying power-

law continuum with varying exponent, intrinsic absorption from gas associated with the quasars, and a plethora of unresolved emission features (e.g. Figure 14).

The challenge is compounded by the fact that no well-developed physical model exists for the quasar SED. The phenomenon is well-modeled by a combination of synchrotron emission, optically thick emission lines arising in a wind, a hot accretion disks emitting preferentially at UV wavelengths, and a partially obscuring dust torus that radiates at IR wavelengths (e.g. Lusso et al, 2012). But each of these is at best empirically parameterized, and may contribute largely independent of one another with great diversity. Furthermore, the emission is poorly constrained at $\lambda_{\text{rest}} < 900\text{\AA}$ in part because the IGM greatly absorbs the SED making it very difficult to assess the intrinsic flux (see Lusso et al, 2015).

Despite all of these vagaries and variabilities, the average quasar spectrum exhibits remarkably little evolution across cosmic time. Figure 15 compares composite quasar SED spectra generated by averaging tens of $z \sim 1$ sources (Telfer et al, 2002) and thousands of $z \sim 2$ sources (Vanden Berk et al, 2001). There are significant differences in the broad emission lines (e.g. Ly α , CIV 1550), but the underlying SED is very similar between the two samples. This commonality holds to the highest redshifts where quasars are observed (Figure 16 Becker et al, 2013). This offers hope that quasar continua can at least empirically be estimated.

Historically, and even today (e.g. O'Meara et al, 2015), the majority of continuum normalization has been performed manually. Redward of Ly α emission, where absorption from intervening gas is minimal (e.g. Figure 14), one can identify unabsorbed regions (by-eye or by-algorithm) to fit the continuum with a model. A higher

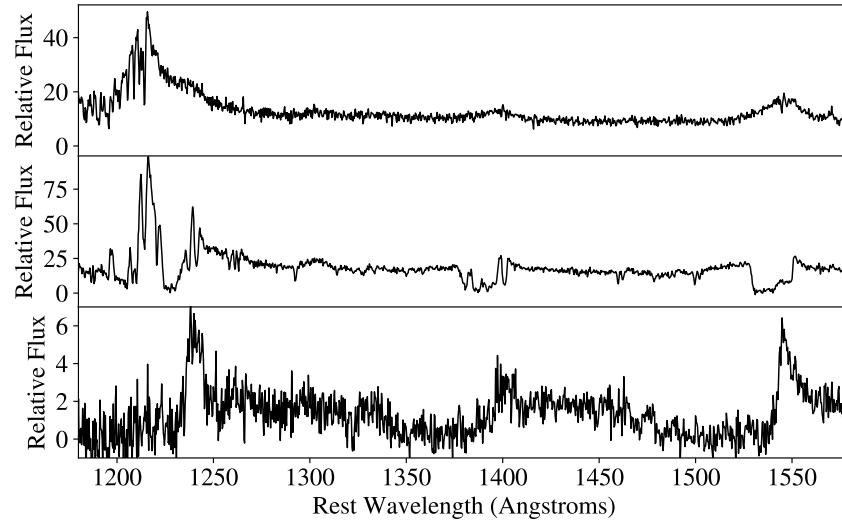


Fig. 14 Three quasar spectra from the SDSS survey (J021740.97–085447.9, J004143.16–005855.0, J003859.34–004252.4) which demonstrate diversity in the degree of associated absorption due to highly ionized gas in the quasar environment.

order ($n \sim 7$) polynomial suffices over each span of $\approx 100\text{\AA}$ to capture undulations due to underlying emission features. One can routinely achieve a precision of several percent, depending on the data quality. An example in the *Spectral Analysis* Notebook shows the call to a GUI used for continuum fitting.

Blueward of Ly α emission, where the IGM opacity is substantial, continuum estimation is very difficult. Human (by-eye) analysis typically involves spline evaluations through spectral regions purportedly free of IGM absorption. But even at echelle resolution, there may be not a single pixel unscathed by the IGM.

Modern analysis has aimed to eliminate the human aspect by empirically predicting the quasar continuum. Suzuki (2006) was the first to perform a Principal Component Analysis (PCA) of quasar continua. This mathematical technique analyzes a cohort of individual spectra (they used the same $z \sim 1$ quasar spectra of the Telfer composite) to calculate the eigenvectors that best describe variations in the spectra off the mean. Formally, we have

$$|q_i\rangle = |\mu\rangle + \sum_{j=1}^m c_{ij} |\xi_j\rangle \quad (64)$$

where $|q_i\rangle$ is any quasar spectrum, $|\mu\rangle$ is the mean (composite) spectrum, $|\xi_j\rangle$ are the Principal Components, and c_{ij} are the eigenvalues. These eigenvectors are orthogonal and generally have little physical meaning. The eigenvectors derived by Suzuki (2006) are shown in Figure 17 and they found the first 3 PCA components accounted for over 80% of the observed variance in the quasar spectra.

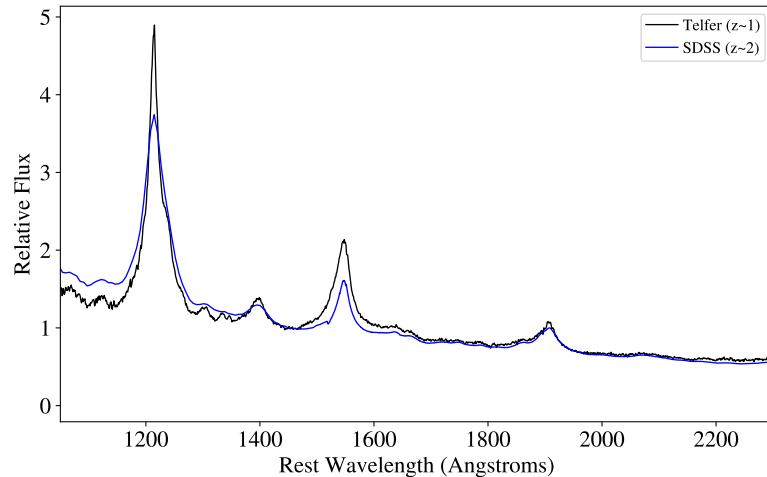


Fig. 15 Comparison of the quasar composite spectrum derived from *HST* spectra of $z \sim 1$ quasars (Telfer et al, 2002) with a composite spectrum derived primarily from $z \sim 2$ quasars taken from the SDSS survey (Vanden Berk et al, 2001). Each composite is normalized to unity at $\lambda_{\text{rest}} \approx 1450\text{\AA}$. While there are notable differences in the strength and widths of the emission line features (dominated by the so-called Baldwin effect), the overall SED is remarkably similar.

The modern approach to quasar fitting, especially in lower resolution spectra, is to fit the quasar SED with these PCA eigenvectors outside the Ly α forest ($\lambda_{\text{rest}} > 1215.67\text{\AA}$) and then extrapolate the model to wavelengths impacted by the IGM. A more recent example of this technique is described in Pâris et al (2011). We also refer the student to analysis that includes Mean Flux Regulation Lee et al (2012) where the PCA estimate is further refined in the Ly α forest by imposing that the IGM opacity match the average (which we derive in a later section).

3.3 Equivalent Width Analysis

Now that we have normalized the spectrum by the intrinsic source continuum, we may proceed to analyze the observed absorption. The fundamental observable of absorption strength is the equivalent width W_λ . Observationally, this metric describes the fraction of incident flux absorbed by the gas. Empirically, there are two standard approaches to estimating W_λ : box-car integration and line-profile analysis.

A non-parametric evaluation may be performed with boxcar integration, i.e., the simple summation over the pixels covering the absorption line. With the flux continuum normalized (and expressed as \bar{f}), we have

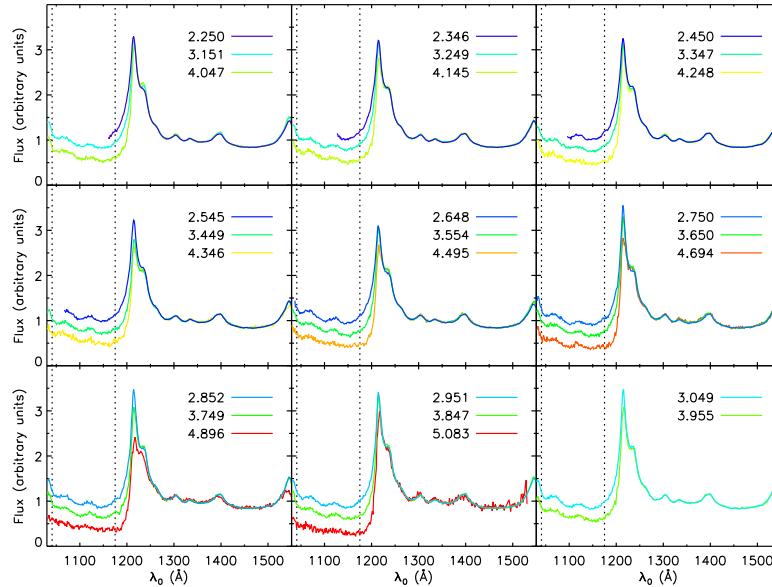


Fig. 16 Composite spectra for quasars drawn from the SDSS over redshifts $z \approx 2.2 - 5$ by Becker et al (2013). Note the closer similarity in the SEDs across these several Gyr. The obvious exception is at $\lambda_{\text{rest}} < 1215\text{\AA}$, where the IGM opacity increase with increasing redshift. We will return to an analysis of the spectra in this figure in the following section.

$$W_\lambda = \sum_i (1 - \bar{f}_i) \delta \lambda_i \quad (65)$$

where $\delta \lambda_i$ is the dispersion of each pixel and the sum is performed over ‘all’ pixels of the profile. The uncertainty follows from propagation of error (neglecting continuum uncertainty),

$$\sigma^2(W_\lambda) = \sum_i \sigma^2(\bar{f}_i) (\delta \lambda_i)^2 \quad (66)$$

One may also adopt parametric techniques by fitting a model of the line-profile to the observed absorption; this is referred to as line-fitting. Most HI absorption features (aside from the very strongest) may be well approximated by a (upside-down) Gaussian. Or by considering $(1 - \bar{f})$ one can fit with

$$g(\lambda) = A \exp[-(\lambda - \lambda_0)^2 / (2\sigma_\lambda^2)] \quad (67)$$

and calculate the equivalent-width as the area under the curve

$$W_\lambda = A \sigma_\lambda \sqrt{2\pi} \quad (68)$$

Uncertainty should include the co-variance between A and σ

$$\sigma(W_\lambda) = W_\lambda \sqrt{\sigma^2(A)/A^2 + \sigma^2(\sigma_\lambda)/\sigma_\lambda^2 + 2\sigma(A)\sigma(\sigma_\lambda)/[A\sigma_\lambda]} \quad (69)$$

We show examples of calculating W_λ in the *Spectral_Analysis* Notebook.

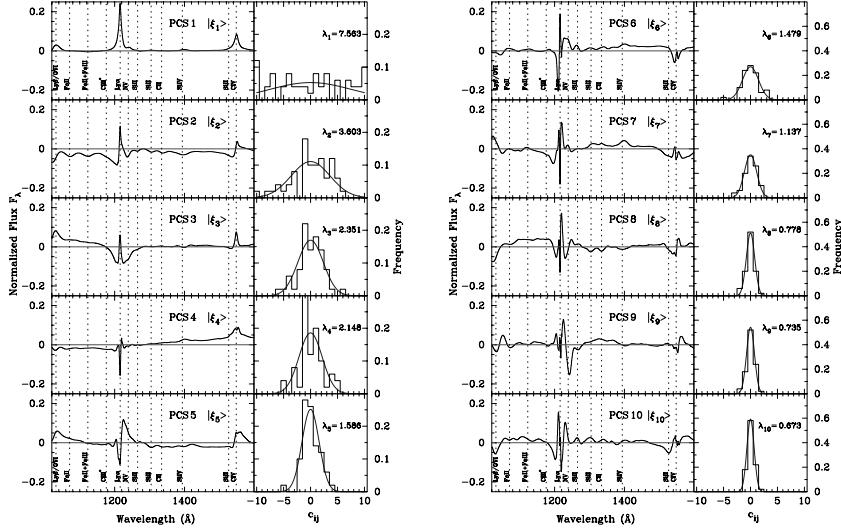


Fig. 17 Each of the sub panels shows one PCA eigenvector ξ_j (left) and the distribution of eigenvalues c_{ij} (right) for the $z \sim 1$ quasars analyzed by Suzuki (2006). The eigenvectors from top-left to lower-right are ordered in decreasing relevance. Taken from Suzuki (2006).

Analysis of W_λ is generally performed in the observer (i.e. measured) frame, but W_λ^{obs} depends on λ and therefore redshift. To relate W_λ to physical quantities (e.g. N, b via a curve-of-growth analysis; Figure 11), one requires a shift to the rest-frame

$$W_\lambda^{\text{rest}} = \frac{W_\lambda^{\text{obs}}}{1+z} \quad (70)$$

When designing an absorption-line observation, one generally must estimate the signal-to-noise (S/N) required to achieve the desired sensitivity. The latter may follow from a desired equivalent width limit W_λ^{lim} . We relate the two quantities by asking (and answering): *What is the limiting equivalent width one can measure from data with a given set of spectral characteristics?* Returning to our boxcar estimate of the uncertainty (Equation 66), we assume a constant dispersion $\delta\lambda$ and recognize that $\sigma(\bar{f})$ is equivalent to the inverse of the S/N per pixel for normalized data. If the S/N is roughly constant across the absorption, then the terms inside the sum in Equation 66 come out and the sum evaluates to M , the number of pixels across the line, i.e.,

$$\sigma(W_\lambda) = \frac{\sqrt{M}\delta\lambda}{S/N} . \quad (71)$$

For a 3σ detection, we relate $W_\lambda^{\text{lim}} = 3\sigma(W_\lambda)$ and we can invert Equation 66 to calculate S/N:

$$S/N = 3 \frac{\sqrt{M}\delta\lambda}{W_\lambda^{\text{lim}}} \quad (72)$$

The lingering unknown is the number of pixels in the integration. If the line is unresolved (or barely resolved), M is simply the sampling ($\Delta\lambda_{\text{FWHM}}/\delta\lambda$). Otherwise, M depends on the intrinsic line-width.

We may also characterize the data quality in terms of a limiting equivalent width. As an example, consider HI Ly α in SDSS spectra (see also the Notebook). For $R = 2,000$, S/N=10 per pixel, and a sampling of 2 pixels, we estimate $W_{\text{lim}}^{3\sigma} \approx 1\text{\AA}$. This falls on the saturated portion of the COG for Ly α (Figure 10). One expects, therefore, to have poor sensitivity to N_{HI} in such spectra. As another example, consider the required S/N for a 30m \AA detection at 3σ significance in an echelle spectrum with $R = 30,000$ and 3 pixel sampling. Using Equation 72, we estimate that a S/N ≈ 10 is desired.

3.4 Line-Profile Analysis

While W_λ offers an intuitive, observationally based measure of the absorption, it is difficult to translate such measurements to physical quantities, i.e. the column density and/or kinematics of the gas. At high spectral resolution ($R > 20,000$), one may fully resolve the line profile and estimate the optical depth directly. In

turn, a line-profile analysis yields the line-centroid (i.e. redshift) and the N, b values described in the previous section. A full discussion of line-profile fitting techniques is beyond the scope of this Chapter. We refer the reader to a simple example in the *Spectral Analysis* Notebook. And we further refer the reader to two of the widely adopted packages for line-profile analysis: VPFIT (developed by R. Carswell): <http://www.ast.cam.ac.uk/~rfc/vpf.fit.html> and ALIS (R. Cooke): <https://github.com/rcooke-ast/ALIS>. Each package performs χ^2 minimization on a set of input absorption lines to derive physical quantities. We will see examples of the results from such analysis in the following sections.

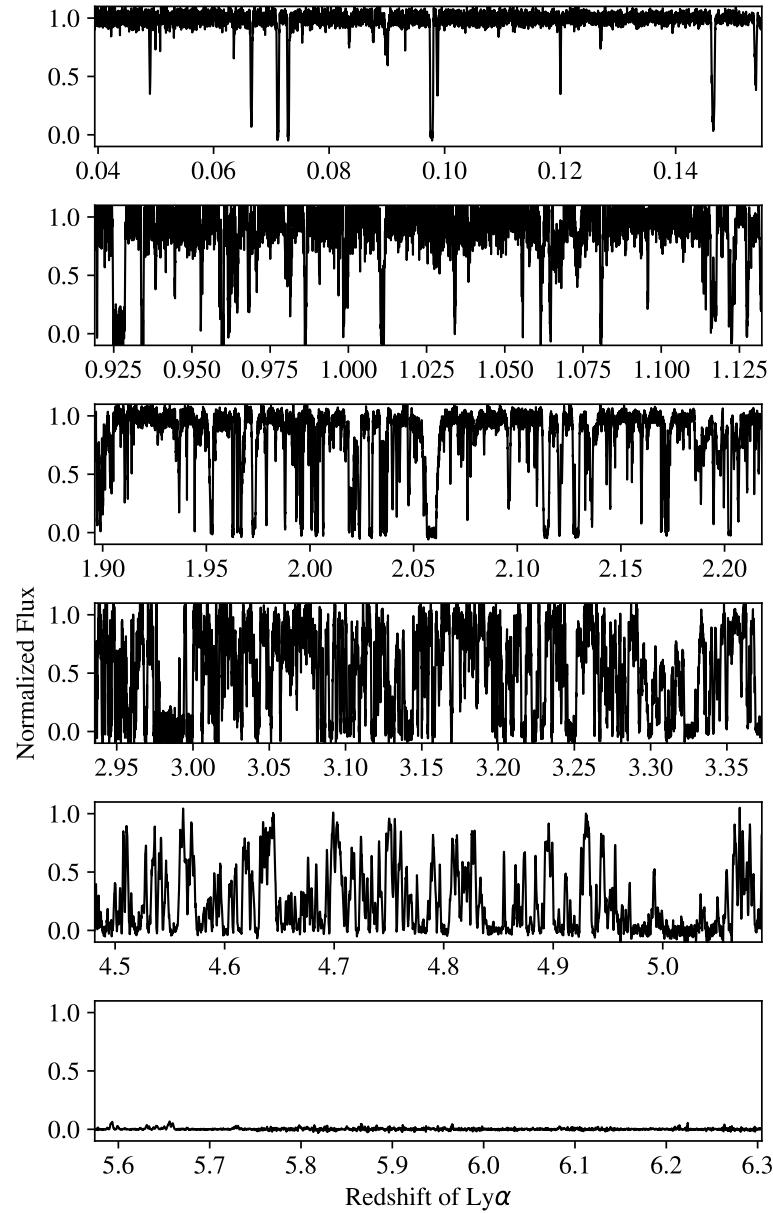


Fig. 18 Snapshots of the Ly α forest in a series of spectra with increasing source redshift. The data are plotted against the redshift for Ly α , i.e., $z_{\text{Ly}\alpha} = \lambda_{\text{obs}}/1215.67\text{\AA} - 1$. Note the evolution from a nearly transparent Ly α forest at $z_{\text{Ly}\alpha} \approx 0$ to essentially zero transmission at $z_{\text{Ly}\alpha} \approx 6$.

4 HI Lines of the Ly α Forest $f(N_{\text{HI}}, z)$

The term Ly α forest refers to the thicket of absorption observed in the spectra of a distant source at $\lambda_{\text{rest}} < 1215.67\text{\AA}$ in the source rest-frame. The absorption lies blueward of the source Ly α emission because these photons redshift while traveling to intervening HI gas at $z < z_{\text{source}}$ to then scatter at Ly α in the rest-frame of this gas. Figure 18 shows Ly α forest spectra for sources ranging from $z \approx 0 - 6$. Two aspects are immediately obvious: (i) the opacity fluctuations are substantial at any given redshift and at least has the appearance of discrete lines; (ii) there is a significant increase in the average opacity with increasing redshift.

In this section, we introduce the standard approaches used to characterize the Ly α forest opacity. These techniques are primarily empirical although the results offer valuable constraints on modern cosmology. Fundamentally, any precise measurement from the IGM offers a valuable constraint on our cosmology because we have a well-developed theory of structure formation (e.g. N-body simulations) and on quasi-linear scales, the baryons track the dark matter. Indeed, one may predict the Ly α opacity provided an estimate of the universe's ambient radiation field (which may also be constrained). This section focuses on results for gas that is optically thin at the HI Lyman limit (i.e. $N_{\text{HI}} < 10^{17} \text{ cm}^{-2}$, as derived in the following section). We note further that true intergalactic gas likely corresponds to even lower column densities ($N_{\text{HI}} < 10^{14} \text{ cm}^{-2}$).

4.1 N_{HI} frequency distribution $f(N_{\text{HI}})$: Concept and Definition

Motivated by the discrete appearance of lines in the Ly α forest (e.g. Figure 18), the first approach developed was to model the gas as a series of absorption lines. Each line has three physical parameters (N, b, z) and the Ly α forest is described by the distributions of N and b across cosmic time. This technique demands echelle resolution spectra and high S/N to precisely constrain the absorption-line parameters.

In practice, one laboriously fits the individual lines in a set of spectra (section 3.4). Figure 19 shows a snippet of spectra analyzed in this manner by Kirkman and Tytler (1997). The approach has the positive features of capturing the stochastic nature of the Ly α Forest and its absorption-line appearance. Furthermore, it is primarily model-independent aside from the underlying ansatz that the gas arises primarily in discrete lines. On the other hand, the N, b, z distributions do not represent a truly physical model and similar quantities are not easily derived from actual models or simulations of the IGM (Rahmati et al, 2013). Furthermore, the analysis tends to be very expensive (laborious), human interaction implies non-reproducible results, and the data required is expensive to obtain (demanding many hours of integration with echelle spectrometers on 10-m class telescopes). Nonetheless, it offers a precise description of the IGM.

Formally, one defines a frequency distribution for the lines of the Ly α forest $f(N_{\text{HI}}, b, z) dN_{\text{HI}} db dz$, the number of lines on average in the intervals $N_{\text{HI}}, N_{\text{HI}} +$

dN_{HI} ; $b, b + db; z, z + dz$. This distribution function is akin to a luminosity function $\phi(L)dL$, which describes the average number of galaxies per volume in a luminosity interval $L, L + dL$. The absorption, however, occurs along sightlines instead of within a volume. A standard assumption adopted in most analyses is that the Doppler parameter (b -value) distribution has minimal N_{HI} (or z) dependence, i.e. the N_{HI} and b -value distributions are separable,

$$f(N_{\text{HI}}, b) = f(N_{\text{HI}}) g(b) \quad (73)$$

Indeed, in the following, we discuss these separately.

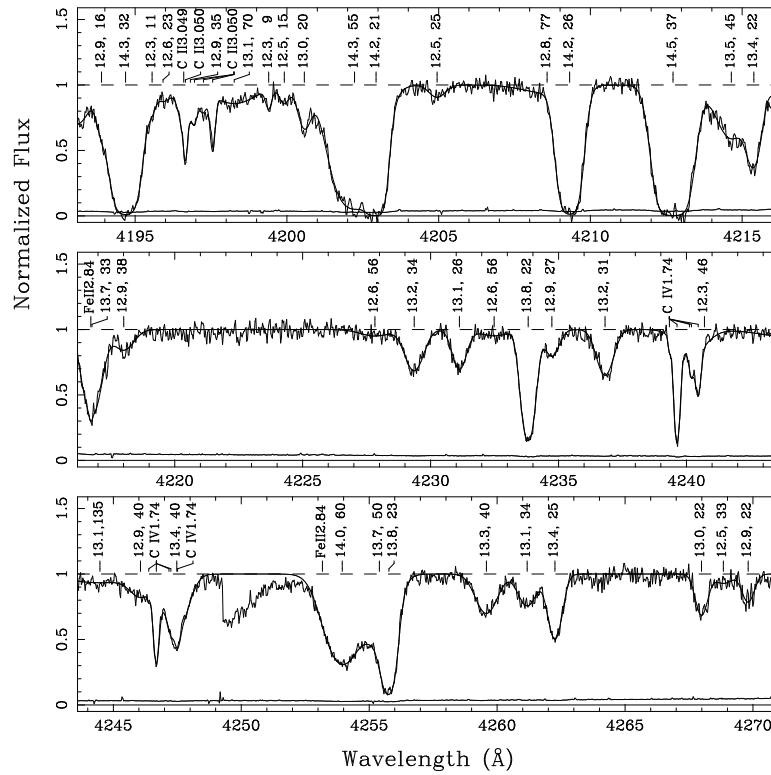


Fig. 19 Examples of a series of fits to lines in high S/N, echelle spectra of the Ly α forest. The dashes indicate individual lines with corresponding N, b values from fits to the data. The redshift of each line is given by the fitted line centroid. Taken from Kirkman and Tytler (1997).

4.2 Binned Evaluations of $f(N_{\text{HI}})$

We may estimate $f(N_{\text{HI}})$ from a set of absorption-line measurements as follows. Consider first, a single sightline to a source at $z = 3$. A spectrum spanning $\lambda = 3500 - 5000\text{\AA}$ will cover Ly α absorption at $z \approx 2 - 3$. One may slice the sightline into multiple redshift intervals, e.g. $\delta z = 0.5$. To bolster the statistics for evaluating $f(N_{\text{HI}})$ at a given redshift, we repeat the experiment for many \mathcal{N} sightlines and cover a survey path (see the *characterizing_lya_forest* Slides for an illustration):

$$\Delta z = \sum_i^{\mathcal{N}} (\delta z)_i \quad (74)$$

Later in this chapter, I will define the ΔX survey path which offers a more physical definition than the observational redshift path.

The simplest estimator of $f(N_{\text{HI}})$ is a binned evaluation of lines having N_{HI} in a finite ΔN_{HI} interval, ideally large enough to have statistical significance. We evaluate,

$$f(N_{\text{HI}}, z) = \frac{\# \text{ lines in } [\Delta N_{\text{HI}}, \Delta z]}{\Delta N_{\text{HI}} \Delta z} \quad (75)$$

where again Δz is the survey path for lines in the ΔN_{HI} interval. Uncertainty is simply assumed to follow Poisson statistics. An important subtlety with this estimator is that evolution in $f(N_{\text{HI}})$ with N_{HI} or z skews the detection within a bin (to lower N_{HI} and higher z). Therefore, simple model fitting with χ^2 analysis using the center of the bins will be (marginally) wrong.

Early results adopting this approach are shown in Figure 20. There is a steep and monotonic decrease in $f(N_{\text{HI}})$ with increasing N_{HI} , partly resulting from the fact that N_{HI} appears in the denominator of Equation 75. As the figures illustrate, over many orders of magnitude in N_{HI} , the data are reasonably well-described with a power-law. This led Tytler (1987) to propose a single-population of gas ‘clouds’. In contrast, Petitjean et al (1993) observed significant departures from a single power-law and argued for distinct populations. Indeed, the modern data and interpretation is even more complex.

With the commissioning of an echelle spectrometer on the 10-m Keck I telescope (Vogt et al, 1994, HIRES), observers had access to exquisite data on multiple sightlines to assess $f(N_{\text{HI}})$. Figure 21 show early results from Kirkman and Tytler (1997) derived from line-profile fitting analysis along the sightline to HS 1946+7658 observed to very high S/N. Their results show the decline in $f(N_{\text{HI}})$ for $N_{\text{HI}} > 10^{12} \text{ cm}^{-2}$ as in previous work, but also an apparent turn-over in $f(N_{\text{HI}})$ at lower N_{HI} . This turn-over at low N_{HI} could reflect a lack of sensitivity (i.e. incompleteness); indeed, HI Ly α with $N_{\text{HI}} = 10^{12} \text{ cm}^{-2}$ has $\tau_0 \approx 0.025$. However, the data are exquisite (see Figure 19) and a careful assessment of incompleteness was performed. Another possibility is that the authors have underestimated the quasar continuum; even a few percent error in the continuum would overwhelm the signal at these optical depths. But the data quality ($S/N > 100$) is such that this is not the most likely explanation. Instead, we are likely witnessing the transition from de-

scribing the Ly α forest by individual lines to a more continuous opacity. We return to this concept towards the end of the section.

Modern estimates of $f(N_{\text{HI}})$ now include tens of sightlines with exquisitely small statistical error (Rudie et al, 2013; Kim et al, 2013). The results, however, are not in uniform agreement suggesting that systematic error is the limiting factor.

4.3 Models for $f(N_{\text{HI}})$

The binned evaluations of $f(N_{\text{HI}})$ describe the shape of distribution function and have motivated empirical models. These models reduce the binned evaluations to a few, well-constrained parameters. In turn, one can examine redshift evolution in the amplitude and shape of $f(N_{\text{HI}})$ and use the model to evaluate $f(N_{\text{HI}})$ at any N_{HI} for calculations on the Ly α forest (see below Meiksin and Madau, 1993). We emphasize, however, that we have no *a priori* physical model for $f(N_{\text{HI}})$ and therefore

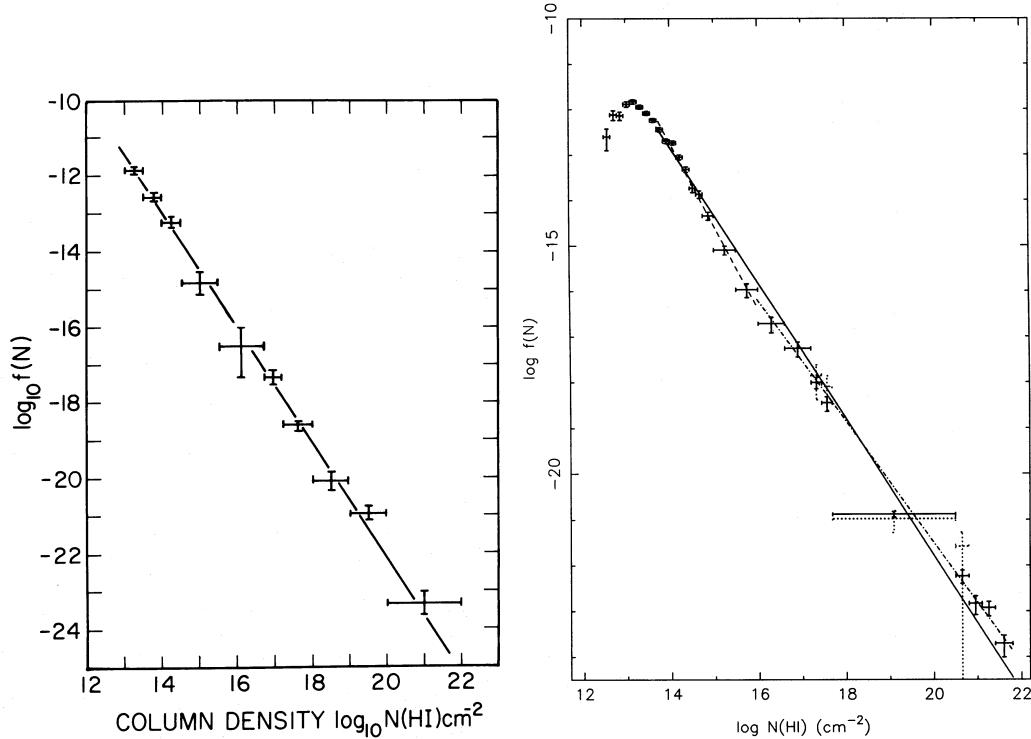


Fig. 20 Binned evaluations of $f(N_{\text{HI}})$ from early data sets on the Ly α forest (left) Tytler (1987), (right) Petitjean et al (1993).

there is the freedom to introduce any functional form. We also note that the results will depend on the N_{HI} range considered.

As the observational constraints have improved, the functional forms adopted for $f(N_{\text{HI}})$ have evolved from a single power-law (Tytler, 1987), $f(N_{\text{HI}}) = BN^{\beta}$ to broken power-laws (Petitjean et al, 1993; Prochaska et al, 2010), to broken, disjoint power-laws (Rudie et al, 2013) to the sum of several functional forms (Inoue et al, 2014), to spline evaluations (Prochaska et al, 2014).

In the early days, simple χ^2 analysis to binned evaluations of $f(N_{\text{HI}})$ was sufficient but these results are sensitive to the choice of bins and did not account for evolution within the bin. A more robust approach is to apply Maximum Likelihood techniques (Storrie-Lombardi et al, 1996; Cooksey et al, 2010).

Consider a sightline sample with a total survey path Δz covering the interval δz at redshift z (refer back to Equation 74). We wish to model $f(N_{\text{HI}})$ at this redshift over a range of N_{HI} . Divide the distribution space into M cells, each with ‘volume’: $\delta V = \delta z \delta N_{\text{HI}}$. Here, we have divided ΔN_{HI} into small intervals and have assumed that $f(N_{\text{HI}})$ does not evolve in the small redshift interval δz considered. The expected number of lines μ_i in i th cell follows from our definition of $f(N_{\text{HI}})$,

$$\mu_i = f(N_{\text{HI}})_i \Delta z \delta N_{\text{HI}} \quad (76)$$

where we consider the full survey path Δz . Therefore, the probability of detecting m absorbers within cell i is given by Poisson statistics,

$$P(m; \mu_i) = e^{-\mu_i} \frac{\mu_i^m}{m!} \quad (77)$$

and construct a Likelihood function as the simple product over all cells

$$\mathcal{L} = \prod_i^M P(m; \mu_i) . \quad (78)$$

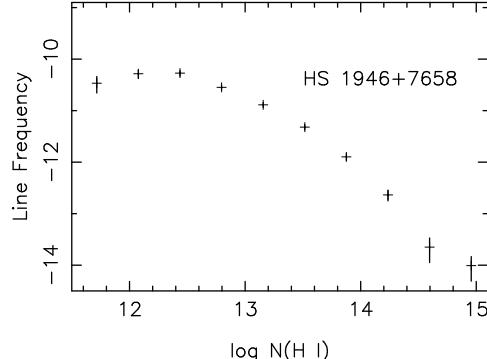


Fig. 21 Estimation of $f(N_{\text{HI}})$ from Keck/HIRES spectroscopy of the very bright quasar HS 1946+7658 by Kirkman and Tytler (1997).

Now perform what appears to be a swindle. Reduce the volume of the cell so that each one contains at most one absorber. This is achieved by letting $\delta N_{\text{HI}} \rightarrow 0$. In this case, $P(m; \mu_i)$ reduces to $e^{-\mu_i}$ for an empty cell and $e^{-\mu_i} \mu_i$ for a cell with 1 system. We may then evaluate the likelihood by summing over all M cells. Explicitly, let there be p lines detected giving $g = M - p$ empty cells. It follows that

$$\mathcal{L} = \prod_{i=1}^g e^{-\mu_i} \prod_{j=1}^p e^{-\mu_j} \mu_j \quad (79)$$

$$= \prod_{i=1}^M e^{-\mu_i} \prod_{j=1}^p \mu_j \quad (80)$$

and the log-Likelihood is expressed as

$$\ln \mathcal{L} = \sum_i^M -\mu_i + \sum_i^p \ln \mu_i \quad (81)$$

$$= \sum_i^M -f(N_{\text{HI}})_i \Delta z \delta N_{\text{HI}} + \sum_j^p \ln f(N_{\text{HI}})_j \Delta z + p \ln [\delta N_{\text{HI}}] \quad (82)$$

In our limit with $\delta N_{\text{HI}} \rightarrow 0$, we ignore the last term (a constant) and take the integral form of the first term to derive

$$\ln \mathcal{L} = - \int_{N_{\min}}^{N_{\max}} f(N_{\text{HI}}) \Delta z dN_{\text{HI}} + \sum_{j=1}^p \ln f(N_{\text{HI}})_j \Delta z \quad (83)$$

If Δz is independent of N_{HI} it may be ignored, but see Cooksey et al (2010) for a treatment where Δz is dependent on N_{HI} .

Lastly, we maximize \mathcal{L} for the parameterization of $f(N_{\text{HI}})$ to derive the “best” values for the parameters. We emphasize that the resultant model need not provide a good description of the data. It is simply the best model for the functional form imposed. The assessment of goodness of fit requires another statistical test (see 5.4).

With modern data and the assumption of a simple power-law, Rudie et al (2013) report $f(N_{\text{HI}}) \propto N_{\text{HI}}^{-1.65 \pm 0.02}$ for $N_{\text{HI}} = 10^{13.5} - 10^{17} \text{ cm}^{-2}$ at $z \approx 2.5$. With a similar but distinct dataset Kim et al (2013) derive $f(N_{\text{HI}}) \propto N_{\text{HI}}^{-1.52 \pm 0.02}$ for $N_{\text{HI}} = 10^{12.75} - 10^{18} \text{ cm}^{-2}$ at $z \approx 2.8$. Formally, these are statistically incompatible. Progress with the $f(N_{\text{HI}})$ approach requires resolving systematic errors, including human biases. Or, alternatively, fundamentally different methods to analyze the data.

4.4 *b*-value Distribution

The results from line-profile fitting also yield a distribution of measured *b*-values. Data from Kirkman and Tytler (1997) are shown in Figure 22. The distribution of *b*-values is dominated by lines with $b \approx 20 - 30$ km/s. Further, one identifies no strong dependence on HI column density aside from a possible increase in the minimum *b*-value with increasing N_{HI} .

Hui and Rutledge (1999) derived a functional form for the *b*-value distribution based on theoretical expectations (fluctuations in the IGM optical depth):

$$g(b) = \frac{4b_\sigma^4}{b^5} \exp\left(-\frac{b_\sigma^4}{b^4}\right) \quad (84)$$

described by a single parameter b_σ . Fitting to the results in Figure 22 they achieved a good description of the observations with $b_\sigma = 26.3$ km/s (Figure 23). This yields an average *b*-value

$$\langle b \rangle = \frac{\int b g(b) db}{\int g(b) db} = b_\sigma \Gamma(3/4) \approx 32 \text{ km/s} \quad (85)$$

It is evident in both of the above figures that the Ly α forest lines exhibits a lower limit to the measured *b*-values. This is $b_{\min} \approx 18$ km/s. What sets this apparent lower bound to the *b*-value (which is well above the spectral resolution of the instrument)?

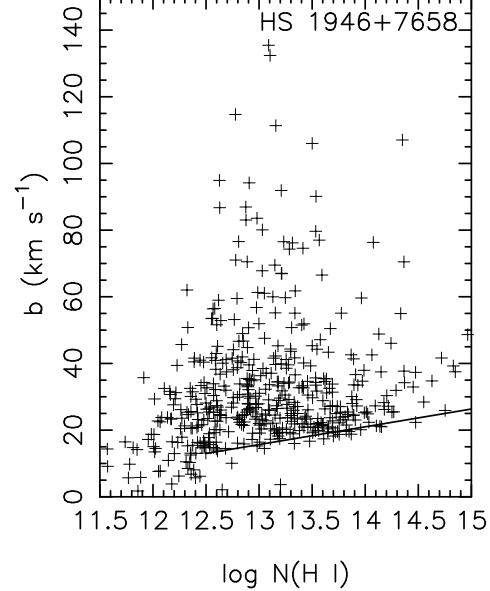


Fig. 22 Scatter plot of *b* vs. N_{HI} measurements for the absorption lines analysed along the sightline to quasar HS 1946+7658 by Kirkman and Tytler (1997).

Recalling that $b = \sqrt{\frac{2kT}{m_A} + \xi^2}$ with ξ^2 a characteristic turbulence of the gas, one is tempted to associate b_{\min} to the IGM temperature. For purely thermal broadening of Hydrogen, we have

$$T = 10^4 \text{ K} \left(\frac{b}{13 \text{ km/s}} \right)^2 \quad (86)$$

and may estimate $T \approx 20,000 \text{ K}$. We comment on recent estimations in the final section.

One also notes an upper bound to the distribution of $b_{\max} \approx 100 \text{ km/s}$. It is difficult to establish whether this represents a sensitivity limit (recall $\tau_0 \propto b^{-1}$) or a true physical limit (e.g. collisional ionization of HI).

4.5 Line Density (the incidence of lines in the Ly α forest)

Referring back to Figure 18, the average opacity of the Ly α forest clearly evolves with redshift. In our description of the IGM as a series of lines, this implies evolution in the line density (or their incidence). Confusingly, the literature is abound with notation for this quantity: $dN/dz, n(z), N(z)$. I have adopted my own: $\ell(z)dz$ defined as the number of lines detected on average in the interval $z, z + dz$ over an interval of column density $N_{\text{HI}} = [N_{\min}, N_{\max}]$. This is the zeroth moment of our frequency distribution,

$$\ell(z)dz = \int_{N_{\min}}^{N_{\max}} f(N_{\text{HI}}, z) dN_{\text{HI}} dz \quad (87)$$

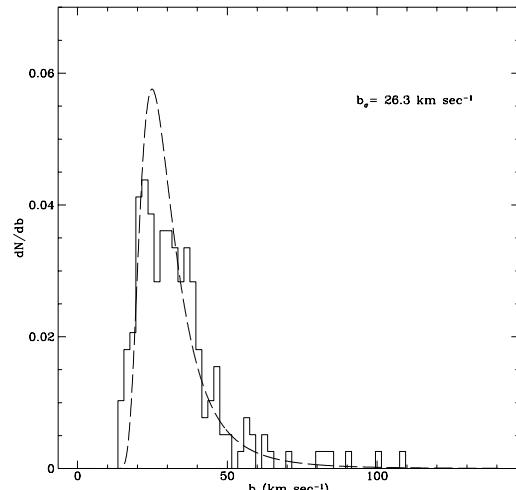


Fig. 23 Fit by Hui and Rutledge (1999) to the observed b -value distribution of lines in the Ly α forest Kirkman and Tytler (1997). This one-parameter model has an average b -value of 32 km/s.

It is akin to the number density of galaxies derived from a luminosity function.

Taking the results for $f(N_{\text{HI}})$ from Kim et al (2013) at $z \approx 2.8$, and integrating from $N_{\text{min}} = 10^{12.75} \text{ cm}^{-2}$ to $N_{\text{max}} = 10^{17} \text{ cm}^{-2}$, we estimate

$$\ell(z \approx 2.8) = \frac{10^{9.1}}{0.52} N_{\text{HI}}^{-0.52} |_{N_{\text{min}}}^{N_{\text{max}}} \approx 560 \quad (88)$$

This integral is dominated, of course, by the lowest N_{HI} systems. If we consider a 5 Å patch of spectrum at 4600 Å, we estimate $\delta z = (1+z)(\delta\lambda/\lambda) = 0.004$ and predict $\mathcal{N} = \ell(z)\delta z = 2.4$ lines on average.

Binned evaluations of $\ell(z)$ from Kim et al (2013) are shown in Figure 24, where one observes strong redshift evolution, as evident in individual spectra. Their analysis also indicates differing behavior for lines of differing N_{HI} implying a likely evolution in the shape of $f(N_{\text{HI}})$.

We can model the redshift evolution in $\ell(z)$, which is also the redshift evolution in the normalization of $f(N_{\text{HI}})$. Here we have some physical guidance on the functional form (see also Meiksin, 2009). Imagine a population of absorbers with physical (proper) number density $n_p(z)$ and proper cross-section $A_p(z)$, e.g. a population

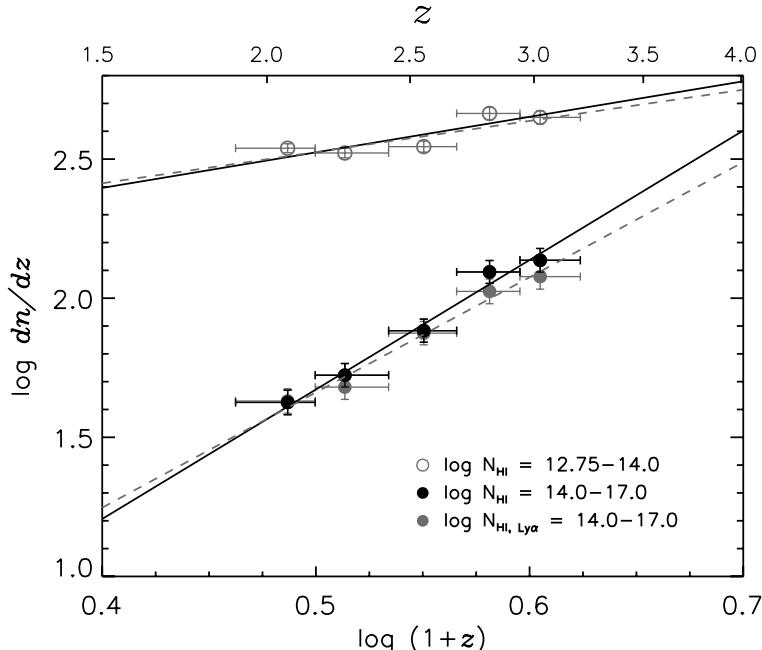


Fig. 24 Evolution in the incidence of Ly α forest lines with redshift for two intervals of N_{HI} . As inferred from a visual inspection of the data (Figure 18), the incidence decreases steeply with decreasing redshift.

of spherical absorbing cows. On average, we expect to intersect $\ell(r) = n_p(z)A_p(z)$ absorbers per proper path length dr_p . In our Cosmology,

$$\frac{dr_p}{dz} = \frac{c}{H(z)(1+z)} . \quad (89)$$

Recognizing $\ell(r)dr = \ell(z)dz$, we have

$$\ell(z) = n_p(z)A_p(z)\frac{c}{H(z)(1+z)} \quad (90)$$

The Hubble Parameter in a flat Λ CDM universe is given by

$$H(z) = H_0\sqrt{\Omega_m(1+z)^3 + \Omega_\Lambda} \quad (91)$$

and at $z > 2$, the universe is matter dominated and Ω_Λ may be ignored. This gives $H(z) \approx H_0\Omega_m^{1/2}(1+z)^{3/2}$

Now ansatz that the absorbing gas (cows) is a constant comoving population, $n_p(z) = n_c(1+z)^3$. This gives

$$\ell(z) \propto n_c(z)A_p(z)(1+z)^{\frac{1}{2}} \quad (92)$$

which (on its own) implies a relatively shallow redshift dependence (indeed, much weaker than observed). However, if the gas is undergoing Hubble expansion, then $\bar{\rho} \propto (1+z)^3$ and one predicts an increasing neutral fraction with z provided the emissivity from ionizing sources does not rise more steeply than $(1+z)^3$. In any case, the number density of ionizing sources should also scale as $(1+z)^\gamma$ and altogether, we have a physical motivation for a $(1+z)^\gamma$ evolution in $\ell(z)$.

Fitting this model to the data gives the results at $z \sim 2 - 3$ shown in Figure 24 which correspond to

$$\ell(z) = 100(1+z)^{1.12 \pm 0.24} \quad (93)$$

for $N_{\text{HI}} < 10^{14} \text{ cm}^{-2}$ and

$$\ell(z) = 0.4(1+z)^{4.14 \pm 0.6} \quad (94)$$

for $N_{\text{HI}} = [10^{14}, 10^{17}] \text{ cm}^{-2}$. Both of these have $\gamma > 1/2$ implying physical evolution in the Ly α forest.

4.6 Mock Spectra of the Ly α Forest

Provided the $f(N_{\text{HI}})$ distribution, one may generate mock spectra of the Ly α forest (with or without an underlying source continuum). This produces an empirical description that basically ignores any underlying physical model and its implications (e.g. clustering of gas in the IGM). However, because it is derived directly from the data it may have a greater realism in many respects.

An example is provided in the *fN* Notebook and the code is available in the *pyigm* package. We provide here the basic recipe:

1. Define a redshift interval δz for the mock Forest
2. Define N_{HI} bounds for $f(N_{\text{HI}})$ (usually the full dynamic range)
3. Calculate the average number of lines ($N_{\text{lines}} = \ell(z)\delta z$)
4. Random draw (Poisson) from N_{lines}
5. Draw N_{HI} values from $f(N_{\text{HI}})$
6. Draw b from $g(b)$
7. Draw z from $\ell(z)$
8. Calculate $\tau_{\lambda,i}$, including all Lyman series lines (as applicable)
 - On a wavelength grid fine enough to capture the line profile
 - i.e. a perfect spectrometer
9. Sum $\tau_{\lambda,i}$
10. Calculate $F_{\lambda} = \exp[-\tau_{\lambda}]$
11. Include spectral characteristics
 - Convolve with instrument LSF
 - Rebin to final wavelength array
 - Add in Noise
12. Multiply in a source SED

4.7 Effective Ly α Opacity: $\tau_{\text{eff}}^{\text{Ly}\alpha}$

As emphasized above, the opacity of the Ly α forest is highly stochastic, both within a given sightline and from sightline to sightline. This implies an underlying, stochastic density field associated with the underling large-scale structure of the universe. While there is great scientific value in the undulations (e.g. Lee et al, 2015), one may also derive valuable constraints and inferences from the mean opacity across cosmic time.

Let $\langle F \rangle_{\text{norm}}$ be the average, normalized flux in the Ly α forest as observed across many sightlines. See the *characterizing_lya_forest* Slides for a visualization. One defines an effective Ly α opacity

$$\tau_{\text{eff},\alpha} \equiv -\ln \langle F \rangle_{\text{norm}} \quad (95)$$

as the average opacity of the IGM from Ly α absorption alone (and generally limited to lower density gas, i.e. $N_{\text{HI}} < 10^{17} \text{ cm}^{-2}$). An alternative description used is the Ly α decrement

$$D_A = 1 - \langle F \rangle_{\text{norm}} \quad (96)$$

Cosmologically (as described below), this average opacity is a balance between the baryon density Ω_b and the ambient radiation field. With an accurate measurement of $\tau_{\text{eff},\alpha}$ one can infer either Ω_b or constrain the radiation field.

In principle, one can estimate $\tau_{\text{eff},\alpha}$ from $f(N_{\text{HI}})$ as described by Møller and Jakobsen (1990); Press et al (1993). We describe the technique here for completeness but note that $\tau_{\text{eff},\alpha}$ is measured directly from observations. As such, it can be used as a constraint on $f(N_{\text{HI}})$. Consider the number of lines \mathcal{N} per unit rest wavelength with

$$d\lambda_{\text{rest}} = \lambda_{\text{rest}} dz / (1+z) . \quad (97)$$

We can calculate the number of lines per $\mathcal{N} = \ell(z)dz$ from our HI frequency distribution

$$\ell(z)dz = \int f(N_{\text{HI}}, b, z)dNdb \quad (98)$$

Translating to a rest wavelength interval,

$$\mathcal{N} = \frac{1+z}{\lambda_{\text{rest}}} \int f(N_{\text{HI}}, b, z)dNdb \quad (99)$$

For these lines, the mean equivalent width is

$$\bar{W}_\lambda = \frac{1+z}{\mathcal{N}\lambda_{\text{rest}}} \int f(N_{\text{HI}}, b, z)W_\lambda(N, b)dNdb \quad (100)$$

If the lines are randomly distributed (i.e. no clustering), then the mean transmission $(1 - D_A)$ is⁵

$$1 - D_A = \exp(-\mathcal{N}\bar{W}_\lambda) \quad (101)$$

An expression for the effective opacity follows,

$$\tau_{\text{eff},\alpha} = \int f(N_{\text{HI}}, b, z)W_\lambda^{\text{Ly}\alpha}(N, b)dNdb . \quad (102)$$

This equation is relatively straightforward, but not analytic. A typical cheat is to assume the average b -value instead of a distribution which gives a one-to-one correspondence between N_{HI} and W_λ . The *fN* Notebook shows example calculations using the *pyigm* software package.

Using $f(N_{\text{HI}})$ from Prochaska et al (2014), we estimate $\tau_{\text{eff},\alpha} = 0.24$ at $z = 2.5$. Figure 25 shows the differential contribution with respect to $\log N_{\text{HI}}$. It is evident that $\tau_{\text{eff},\alpha}$ is dominated by lines with $\tau_0 \approx 1$ and that the result depends on the choice of N_{min} .

Returning to actual measurements of $\tau_{\text{eff},\alpha}$, the first results were derived from 24 quasar spectra observed at Lick Observatory (Kirkman et al, 2005). The authors employed an ‘army’ of undergrad students to fit continua, normalize the data, and assess systematic uncertainty. After masking metal absorption and HI lines with $N_{\text{HI}} > 10^{17} \text{ cm}^{-2}$, they reported D_A measurements for $z \approx 2 - 3$ (Figure 26). Mod-

⁵ This equation appears intuitive yet a proper derivation is remarkably complex!

eling D_A as $(1+z)^\gamma$, they found

$$D_A = 0.0062(1+z)^{2.75} \quad (103)$$

which gives $\tau_{\text{eff},\alpha} = 0.22$ at $z = 2.5$.

Analyzing a larger dataset ($\mathcal{N} \sim 100$) at higher spectral resolution, Faucher-Giguère et al (2008b) extended the results to $z = 4$ and improved the precision of the measurements. These authors corrected statistically for metal absorption, assessed continuum bias on mock spectra, and reported on a disturbing ‘wiggle’ in their measurements (which has appeared to disappear in later datasets). Their best-fit power law is $\tau_{\text{eff},\alpha} = 0.0018(1+z)^{3.92}$.

Most recently, Becker et al (2013) have considered the full dataset of SDSS and leveraged the nearly constant mean continuum of quasar spectra (see Figure 16) to measure $\tau_{\text{eff},\alpha}$ in stacked, composite spectra. Their analysis yields the *relative* evolution in $\tau_{\text{eff},\alpha}$ with redshift (Figure 27). Tying their relative measurement to the absolute value at $z = 2.5$ in Faucher-Giguère et al (2008b), they report

$$\tau_{\text{eff},\alpha}(z) = 0.751 \left(\frac{1+z}{1+3.5} \right)^{2.9} - 0.132 \quad (104)$$

which includes statistical corrections for $N_{\text{HI}} > 10^{17} \text{ cm}^{-2}$ absorbers and metal absorption. These measurements provide a blunt yet powerful test for any cosmological model of the IGM.

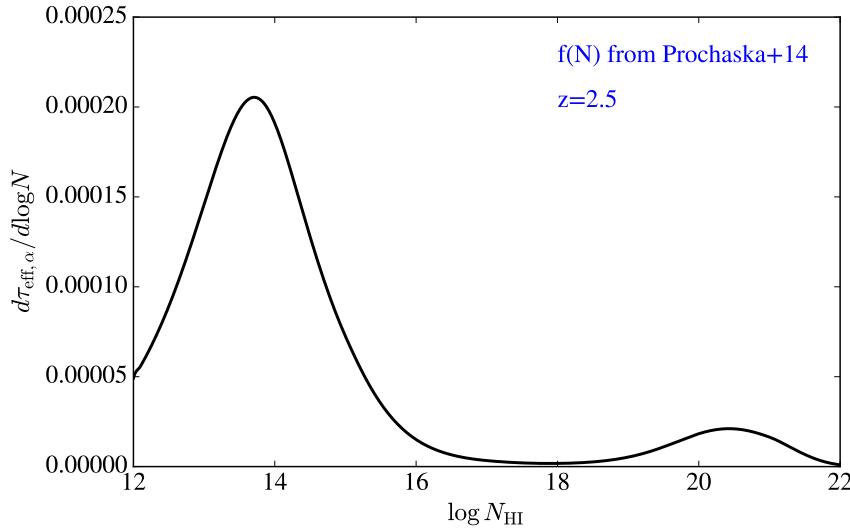


Fig. 25 Differential contribution to $\tau_{\text{eff},\alpha}$ for $z = 2.5$ using the $f(N_{\text{HI}})$ function from Prochaska et al (2014).

4.8 Fluctuating Gunn-Peterson Approximation (FGPA)

The modern paradigm of the Ly α forest, or IGM, is that the observed opacity results from undulations in the underlying dark matter density. The baryons, which are predominantly ionized with an ionization state dictated by photoionization from the extragalactic UV background (EUVB), trace these dark matter fluctuations to yield the observed Ly α forest.

This paradigm lends to the formalism now known as the fluctuating Gunn-Peterson approximation (FPGA). The standard formalism is derived as follows. We begin with our result from quantum mechanics that the optical depth of Ly α at line center is

$$\tau_0 = \frac{\pi e^2}{m_e c} \frac{N_{\text{HI}} f_{\text{Ly}\alpha}}{\Delta v_D} . \quad (105)$$

Expressing the N_{HI} column density as

$$N_{\text{HI}} = n_{\text{HI}} \Delta r \quad (106)$$

with Δr an interval in space and n_{HI} the gas density. From Cosmology (see Equation 89 for dr/dz), we can relate distance to redshift

$$\Delta r = \frac{c \Delta z}{H(z)(1+z)} , \quad (107)$$

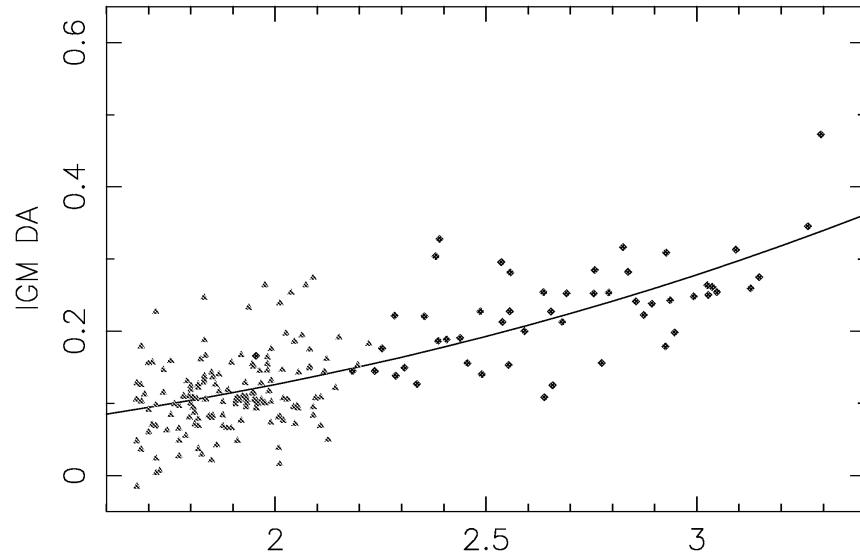


Fig. 26 Measurements of the Ly α decrement D_A at $z \approx 2 - 3$ from a modest sample of quasar spectra, by Kirkman et al (2005). The fitted line is given by Equation 103.

and we have

$$\tau_0 = \frac{\pi e^2}{m_e} \frac{n_{\text{HI}} \Delta z f_{\text{Ly}\alpha}}{\Delta v_D H(z)(1+z)} \quad (108)$$

Because the absorption occurs over a small redshift interval, we can express $\Delta z/(1+z) = \Delta v_{\text{Ly}\alpha}/v_{\text{Ly}\alpha}$ and identify Δv_D in Equation 108 with $\Delta v_{\text{Ly}\alpha}$ (i.e. the line width is given by cosmic expansion). Ionization balance for n_{HI} with ionization rate Γ and recombination rate $\alpha(T)$ gives

$$n_{\text{HI}} \Gamma = \alpha(T) n_{\text{HII}} n_e \quad (109)$$

Altogether now,

$$\tau = \frac{\pi e^2 f_{\text{Ly}\alpha}}{m_e v_{\text{Ly}\alpha}} \frac{1}{H(z)} \frac{\alpha(T) n_{\text{HII}} n_e}{\Gamma} \quad (110)$$

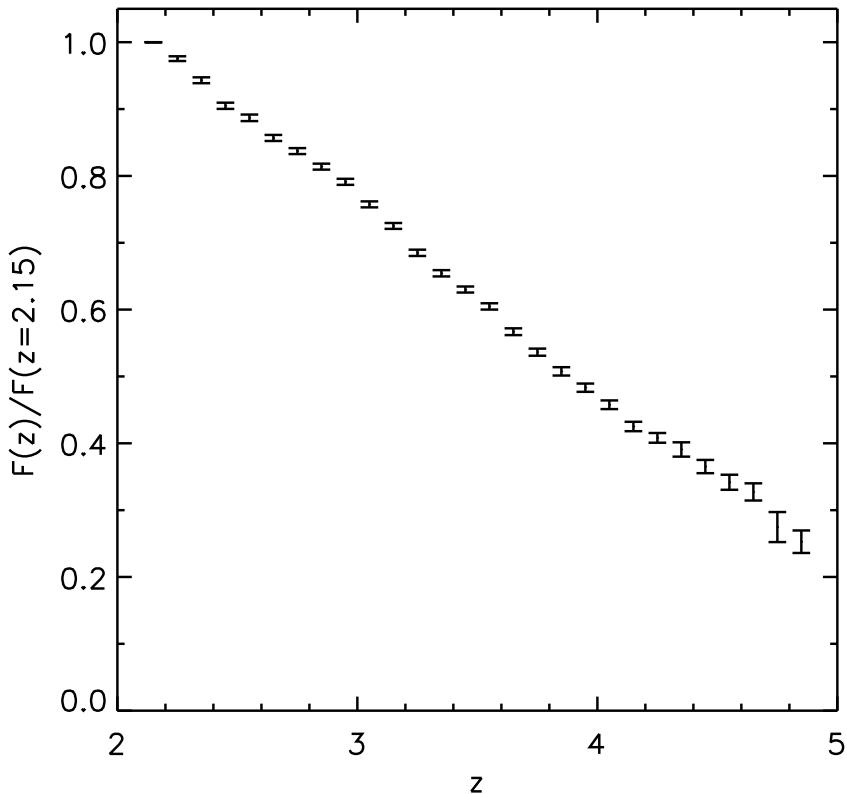


Fig. 27 Relative measurements of the mean flux in the Ly α forest as measured in composite quasar spectra by Becker et al (2013). After tying their measurements to the $z \approx 2.5$ effective optical depth estimate of Faucher-Giguère et al (2008b), they constructed a model for the evolution in $\tau_{\text{eff},\alpha}$ given by Equation 104.

To link the gas density with the dark matter density, we introduce the over-density δ where

$$\rho = \bar{\rho}(1 + \delta) \quad (111)$$

with ρ and $\bar{\rho}$ the dark matter density and its cosmic mean. We then express our number densities in terms of the over-density. Furthermore, we define ionization and mass fractions – X is the Hydrogen mass fraction and x is the ionized Hydrogen fraction – to express

$$n_{\text{HII}} = \frac{\rho_{\text{crit}} \Omega_b}{m_p} Xx(1 + \delta)(1 + z)^3 \quad (112)$$

A similar (uglier) expression for n_e includes Helium (with Y, y_{II}, y_{III} defined similarly):

$$n_e = \frac{\rho_{\text{crit}} \Omega_b}{m_p} [Xx + 0.25Y(y_{II} + 2y_{III})] (1 + \delta)(1 + z)^3 \quad (113)$$

For gas with $T \sim 10^4 \text{ K}$ (as expected for a photoionized medium), $\alpha(T)$ is a power-law of the form,

$$\alpha(T) \approx \alpha_0 T^{-0.7} . \quad (114)$$

Lastly, one assumes the IGM gas follows a power-law temperature-density relation, (as derived from hydrodynamic analysis by Hui and Gnedin (1997))

$$T = T_0(1 + \delta)^\beta \quad (115)$$

The FGPA expression becomes

$$\tau = A(z)(1 + \delta)^{2-0.7\beta} \quad (116)$$

where the opacity “fluctuates” with the local over-density. The $A(z)$ term is proportional to $(1 + z)^6/[H(z)\Gamma]$. or, in its glory,

$$A(z) \equiv \frac{\pi e^2 f_{\text{Ly}\alpha}}{m_e v_{\text{Ly}\alpha}} \left(\frac{\rho_{\text{crit}} \Omega_b}{m_p} \right)^2 \frac{1}{H(z)} Xx[Xx + 0.25Y(y_{II} + 2y_{III})] \frac{\alpha_0 T_0^{-0.7}}{\Gamma} (1 + z)^6 \quad (117)$$

Given a probability density function for the overdensity $P(\Delta)$ with $\Delta \equiv 1 + \delta$, we may calculate the mean flux (and opacity)

$$\langle F \rangle(z) = \int_0^\infty P(\Delta; z) \exp(-\tau) d\Delta . \quad (118)$$

See Miralda-Escudé et al (2000) for models of $P(\Delta)$ based on numerical simulations. This FGPA provides an analytic expression, calibrated against cosmological simulations, for the IGM opacity.

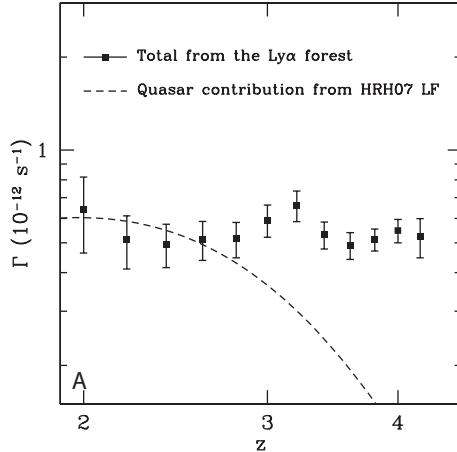


Fig. 28 Estimates of the ionization rate Γ from measurements of the effective optical depth in Ly α and adopting the fluctuating Gunn-Peterson approximation. One recovers a nearly constant value at $z \approx 2 - 4$ in contrast with the estimation from quasars alone over that same epoch (which may decline steeply at high z). Taken from Faucher-Giguère et al (2008a) and HRH07 (?).

Returning to $\tau_{\text{eff},\alpha}$ measurements, we can relate Equation 118 to the data with the only significant unknown⁶ being the photoionization rate Γ . Figure 28 from Faucher-Giguère et al (2008a) shows estimates for Γ based on the $\tau_{\text{eff},\alpha}$ measurements of Faucher-Giguère et al (2008b). The figure also shows their estimate for Γ from quasars alone and the authors argued that star-forming galaxies must dominate the radiation field by $z \sim 3$. Recent work, however, has raised the possibility that faint quasars play as strong a role (Giallongo et al, 2015; Madau and Haardt, 2015). This issue awaits deeper, high- z surveys of faint AGN.

4.9 Effective Lyman Series Opacity

We finish this section by extending our calculations for the effective opacity of Ly α to the full HI Lyman series. Each line up the HI energy ladder contributes additional opacity at shorter wavelengths with sequentially smaller opacity ($\tau_0 \propto \lambda fN$). A complex ‘blending’ of the Lyman series from various redshifts occurs.

For example consider the quasar light emitted at $\lambda_{\text{rest}} = 920\text{\AA}$ from a quasar at $z_{\text{em}} = 4$ (we observe these photons at $\lambda_{\text{obs}} = 4600\text{\AA}$). The photons emitted by the quasar will redshift into resonance with HI Ly α at $z_{\text{Ly}\alpha} = 2.78$, Ly β at $z_{\text{Ly}\beta} = 3.48$, Ly γ at $z_{\text{Ly}\gamma} = 3.73$, Ly δ at $z_{\text{Ly}\delta} = 3.84$, etc.

The effective opacity from each Lyman series line is independent and sums simply. An example calculation for the predicted transmission \mathcal{T} through the IGM with $\mathcal{T} = \exp(-\tau_{\text{eff}})$ is shown in Figure 29. This ‘sawtooth’ curve has teeth at each of the additional Lyman series lines. The slope of increasing transmission with decreasing wavelength is due to the decreasing incidence of lines with decreasing redshift. The f/N Notebook shows an example calculation using software from *pyigm*.

⁶ There is a weak dependence on T which leads to some degeneracy.

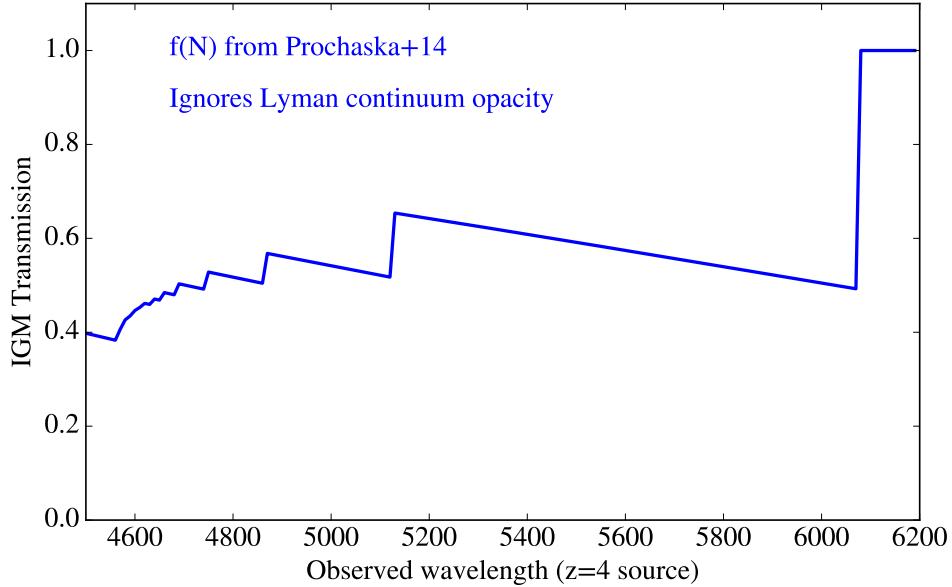


Fig. 29 The so-called ‘sawtooth’ transmission function \mathcal{T} of the HI Lyman series of the IGM for a source at $z = 5$. Redward of HI Ly α there is no attenuation ($\mathcal{T} = 1$). From $\lambda_{\text{obs}} \approx 5100 - 6000\text{\AA}$, only HI Ly α contributes and the increasing transmission with decreasing redshift is due to evolution in the Ly α line density. At shorter wavelengths, additional HI Lyman series lines contribute and reduce the transmission accordingly.

In Figure 30, we show the average relative flux (normalized to unity at $\lambda_{\text{rest}} = 1450\text{\AA}$) of ≈ 150 , $z \sim 4$ quasars from the SDSS. Overlayed on the data is the quasar SED from Telfer et al (2002) attenuated by the effective opacity shown in Figure 29. The good agreement down to $\lambda \approx 4700\text{\AA}$ is truly remarkable and demonstrates the quality of our characterization of the IGM.

Of course, IGM attenuation is imprinted in the spectra of *all* distant sources. This includes galaxies and for a $z = 4$ galaxy, the g -band flux suffers a decrement of $\Delta m_g \approx 1$ mag. This insight led to the discovery of the first high- z star-forming galaxies (Madau, 1995; Steidel et al, 1996; Lowenthal et al, 1997). This attenuation will also affect our Ly α emitters (LAEs). At $z = 5$, $\tau_{\text{eff},\alpha} \approx 1.5$ which likely contributes to the observed asymmetry in Ly α emission.

5 Optically Thick HI Absorption

Thus far, our discussion has focused entirely on Lyman series opacity for atomic Hydrogen (i.e. Ly α , Ly β). These are the bound-bound transitions of HI. The atom also exhibits a bound-free continuum opacity beyond the so-called Lyman limit at

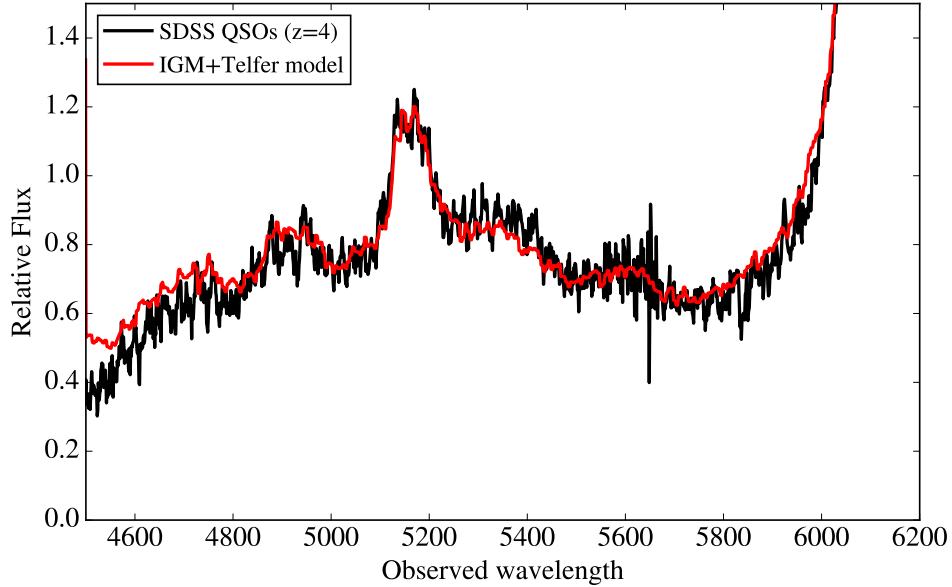


Fig. 30 Composite spectrum of 150 quasars at $z \sim 4$ from the SDSS (black). These are shown in the observer frame for $z = 4$ (but were stacked in the rest-frame) and were normalized to unity at $\lambda_{\text{rest}} = 1450\text{\AA}$. Overlayed on the data is a model using the quasar SED from Telfer et al (2002) and our estimate of the IGM transmission from the full HI Lyman series. The excellent agreement (down to at least 4700\AA) is quite remarkable.

energies $h\nu \geq I_H$ with I_H the ionization energy (13.6eV). This corresponds to wavelengths $\lambda < 911.7\text{\AA}$. The HI continuum opacity leads to additional attenuation of the radiation field at high z , both for ambient radiation (frequently referred to as the extragalactic UV background or EUVB) and for individual sources.

Theoretically, one predicts that optically thick gas originates in environments within and around galaxies and, possibly, denser gas within the IGM. Work on the gas around galaxies (also referred to as the circumgalactic medium or CGM) indicates a high incidence of such gas to at least 100 kpc (Bergeron and Boisse, 1991; Prochaska et al, 2013b, 2017). In turn, the study of optically thick HI bears on the accretion of gas onto galaxies (aka cold flows) and the flow of cool gas away from galaxies (aka feedback).

In this section, we focus on optically thick HI absorption, commonly referred to as Lyman Limit Systems (or LLSs). We also provide a short subsection on the damped Ly α systems (DLAs) which are systems with the highest HI column densities.

5.1 Physics of Continuum Opacity

The bound-free absorption process occurs when a photon with energy $\hbar\omega > I_H$ strikes an atom to eject an electron with energy $\hbar\omega - I_H > 0$, i.e. absorption occurs over a continuous range of frequencies. Our approach to describing the physics of this encounter is to combine an expression for the probability of absorption (w_{fi}) with the density of final states and then calculate the probability/time for a transition to a ‘cell’ of phase space. We will express this probability in terms of a cross-section which will then (when combined with N_{HI}) provide the optical depth.

Consider the final state for the emitted, free electron

$$\hbar\mathbf{k}_f = \mathbf{p}_f \quad (119)$$

and assume a wave function normalized to a very large volume L^3

$$|\mathbf{k}_f\rangle \equiv \frac{1}{L^{3/2}} |f\rangle \quad (120)$$

with energy

$$E_f = \frac{p_f^2}{2m} = \frac{\hbar^2 k_f^2}{2m} \quad (121)$$

The density of final states using a usual phase-space argument (i.e. periodic boundary conditions) is

$$k_x = \frac{2\pi n_x}{L} \quad \text{with } n_x \text{ any integer} \quad (122)$$

For final momentum $\hbar k_f$ propagating in direction $d\Omega$, the density of final states is

$$n^2 dnd\Omega \quad \text{with } n^2 = n_x^2 + n_y^2 + n_z^2 = k_f^2 \left(\frac{L}{2\pi} \right)^2. \quad (123)$$

Finally, we derive $g(E)$, the density of final states, by equating

$$\begin{aligned} n^2 dnd\Omega &= g(E) dEd\Omega \\ &= \left(\frac{L}{2\pi} \right)^3 \frac{m_e}{\hbar^2} k_f dE d\Omega \end{aligned} \quad (124)$$

We may define the Differential Cross-section as

$$d\sigma = \frac{(\text{Energy/unit time}) \text{ absorbed by atom}(i \rightarrow f)}{\text{Energy flux of the radiation field}} \quad (125)$$

The transition probability per unit time (without derivation) is given by

$$w_{fi} = \frac{4\pi^2 e^2}{m^2 c} \frac{I(\omega_{fi})}{\omega_{fi}^2} \left| \langle \phi_f | e^{i\mathbf{k}\cdot\mathbf{r}} \hat{\epsilon} \cdot \nabla | \phi_i \rangle \right|^2 \quad (126)$$

For our photoelectric effect,

$$d\sigma_p = \frac{w_{fi}g(E)d\Omega \cdot \hbar\omega}{I(\omega)} \quad (127)$$

Some algebraic manipulation gives

$$\frac{d\sigma_p}{d\Omega} = \frac{\alpha}{2\pi} \frac{k_f}{m\hbar\omega} |< f | e^{i\mathbf{k}\cdot\mathbf{r}} \hat{\mathbf{e}} \cdot \mathbf{p} | \phi_i >|^2 \quad (128)$$

Finally, we adopt the dipole approximation, which is valid for energies near the ionization potential ($\hbar\omega \approx I_H$). In this case, instead of $|f> = e^{i\mathbf{k}_f \cdot \mathbf{r}}$, we let $e^{i\mathbf{k} \cdot \mathbf{r}} = 1$ assuming $ka_0 \ll 1$, which is automatically satisfied near the threshold energy. The differential cross-section becomes

$$\frac{d\sigma_p^D}{d\Omega} = \frac{\alpha}{2\pi} \frac{m\omega k_f}{\hbar} (\hat{\mathbf{e}} \cdot \mathbf{r})_{fi} \quad (129)$$

For photoionization out of the ground state (1s),

$$|i> = |n = 1; \ell = 0; m = 0; m_s = \pm > \quad (130)$$

The total cross-section, integrating over $d\Omega$, is evaluated as,

$$\sigma_{photo}^D = \frac{2^9 \pi^2}{3} \alpha a_0^2 \left(\frac{I_H}{\hbar\omega} \right)^4 f(\eta) \quad (131)$$

where

$$f(\eta) \equiv \frac{\exp[-4\eta \cot^{-1} \eta]}{1 - \exp[-2\pi\eta]} \quad (132)$$

$$\eta \equiv \left(\frac{I_H}{E_f} \right)^{\frac{1}{2}} = \left(\frac{I_H}{\hbar\omega - I_H} \right)^{\frac{1}{2}} \quad (133)$$

For general atomic K-shell (n=1) absorption, replace

$$\begin{aligned} I_H &\rightarrow Z^2 I_H \\ a_0 &\rightarrow a_0 Z \\ \eta &\rightarrow [Z^2 I_H / (\hbar\omega - Z^2 I_H)]^{\frac{1}{2}} \\ \alpha &\rightarrow 2\alpha \quad \{2 \text{ K-shell electrons}\} \end{aligned}$$

Near threshold, $\hbar\omega \approx I_H$ giving $\eta \gg 1$ and we can approximate

$$f(\eta) \approx e^{-4+4/3\eta^2} \approx e^{-4} \left(1 + \frac{4}{3\eta^2} \right) \quad (134)$$

$$\approx e^{-4} \left(1 + \frac{1}{\eta^2} \right)^{4/3} = e^{-4} \left(\frac{\hbar\omega}{I_H} \right)^{4/3} \quad (135)$$

Altogether, one finds the following frequency dependence:

$$\sigma \propto \begin{cases} 0 & \hbar\omega < I_H \\ v^{-8/3} & \hbar\omega \approx I_H \\ v^{-3} & \hbar\omega \gtrsim I_H \\ v^{-7/2} & \hbar\omega \gg I_H \end{cases} \quad (136)$$

See the *Optically_Thick* Notebook for an illustration. As an aside, one can use the Milne relation to relate the photoionization cross-section to the recombination rate.

Evaluating Equation 131 at $h\nu \approx 1$ Ryd, we calculate

$$\sigma_{\text{photo}}(1 \text{ Ryd}) = 6.339 \times 10^{-18} \text{ cm}^2 \quad (137)$$

and ask: What HI column density is required to give $\tau_{\text{LL}} = 1$? Of course, our continuum opacity is simply $\tau_{\text{LL}}(v) = \sigma_{\text{phot}}(v)N_{\text{HI}}$ which we can invert to give an expression for column density:

$$N_{\text{HI}} = \frac{\tau_{\text{LL}}}{\sigma_{\text{phot}}} \quad (138)$$

Evaluating at 1 Ryd, we estimate $N_{\text{HI}} = 10^{17.198} \text{ cm}^{-2}$. Therefore, an HI gas with integrated $N_{\text{HI}} = 10^{17.2} \text{ cm}^{-2}$ will exhibit a Lyman limit optical depth of unity at $\lambda_{\text{LL}} = 911.7 \text{ \AA}$. This opacity is continuous and declines as $v^{-\alpha}$ with $\alpha \approx 3$ (closer to 2.75).

5.2 Lyman Limit System (LLS)

In quasar absorption line parlance, we define an HI absorption system that is optically thick at the Lyman limit to be a Lyman Limit System or LLS. The value of τ_{LL} defining an LLS is somewhat arbitrary, but $\tau_{\text{LL}} \geq 1$ is probably the most sensible⁷. Given equation 138, this implies HI systems with $N_{\text{HI}} \geq 10^{17.198} \text{ cm}^{-2}$ or $\log N_{\text{HI}} \geq 17.2$. At this column density, the first ≈ 10 lines of the Lyman series are highly saturated ($\tau_0 \gg 1$). Furthermore, The integrated opacity of the highest order Lyman series lines matches τ_{LL} at 1 Ryd, i.e. the effective optical depth is continuous through the Limit. This is illustrated in Figure 31.

⁷ . With an LLS defined as HI gas with $\tau_{\text{LL}} \geq 1$, systems with $\tau_{\text{LL}} \lesssim 1$ are termed partial LLS or pLLS.

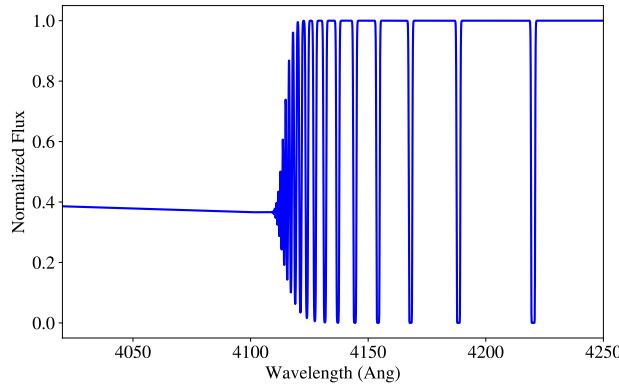


Fig. 31 Idealized absorption spectrum of an LLS at $z = 3.5$ with $N_{\text{HI}} = 10^{17.2} \text{ cm}^{-2}$. One sees the Lyman series absorption at $\lambda > 4110\text{\AA}$ that converges to the continuum opacity at $\lambda_{\text{rest}} < 911.7\text{\AA}$ with optical depth of 1.

Figure 32 shows a real-world example, the LLS identified in the VLT/X-Shooter spectrum of J0529–3526 taken from the XQ-100 Survey (López et al, 2016). The LLS at $z \approx 4.37$ ($\lambda_{\text{LL}} \approx 4900\text{\AA}$) absorbs approx 66% of the quasar flux that is transmitting through the Ly α forest. This implies an optical depth from the continuum opacity of $\tau_{\text{LL}} \approx 1$ and therefore $N_{\text{HI}} \approx 10^{17.2} \text{ cm}^{-2}$. Careful inspection reveals a second LLS at $\lambda \approx 4470\text{\AA}$ which completely absorbs the quasar flux (i.e. $\tau_{\text{LL}} > 3$).

In analysis of actual spectra (e.g. Figure 32), one can recover a precise measurement for the optical depth only over a limited range: $\tau_{\text{LL}} = 0.2 – 3$. For $\tau > 3$ the flux is zero and one only establishes a lower limit to the opacity. For $\tau < 0.2$, the flux decrement is difficult to detect amidst the fluctuating IGM and variations in the quasar SED. This limited range of τ_{LLS} implies a limited range of sensitivity for measuring N_{HI} from the Lyman limit of $N_{\text{HI}} \approx 10^{16.5} – 10^{17.8} \text{ cm}^{-2}$.

5.3 LLS Surveys

As a first estimate for the universe's average opacity to ionizing photons, one may survey for LLS in a large spectroscopic sample of sources. This requires, of course, spectra blueward of the Lyman Limit and that the source emits photons beyond its Lyman limit. Conveniently, most quasars are transmissive at $h\nu > 1$ Ryd as illustrated by the *HST/WFC3* quasar spectrum shown in Figure 33. Evidently, Type I Quasars are sufficiently bright to photoionize the sightline through their own host galaxy. This proximity effect extends to 1 Mpc and possibly beyond (Prochaska et al, 2013a).

A survey for LLS is relatively straightforward: search for and measure a continuum break associated to the Lyman continuum opacity. There is, however, a subtlety. The discovery of one LLS generally precludes the discovery of any others at lower

redshift because the quasar flux has been significantly attenuated. This is unique amongst absorption systems or any other astrophysical phenomena. And it inspires a unique approach to analysis of the survey.

Tytler (1982) introduced the concepts and techniques of Survival Statistics to analyze his own survey of LLS. We briefly present his formalism. Let t be the redshift separation of the LLS from the QSO redshift, i.e. $t = z_{\text{em}} - z_{\text{LLS}}$. For a spectrum with finite spectral coverage (λ_{\min} to λ_{\max}), the LLS search must truncate at $z_{\min} = \lambda_{\min}/\lambda_{\text{LL}} - 1$. For example, $z_{\min} \approx 3.2$ for $\lambda_{\min} = 3800\text{\AA}$ for spectra from the SDSS survey. The maximum value of t for a given sightline is therefore $T = z_{\text{em}} - z_{\min}$.

Define three functions to describe the distribution of LLS: (1) the Observed density function, the probability that an LLS is observed in a line of sight at $t = x$

$$f(t) = \frac{\text{Prob}(t < x < t + \Delta t)}{\Delta t} \text{ as } \Delta t \rightarrow 0 \quad (139)$$

(2) the Survival function, the probability that *no* LLS occurs in the interval $(0, t)$

$$S(t) = 1 - \int_0^t f(t') dt' \quad (140)$$

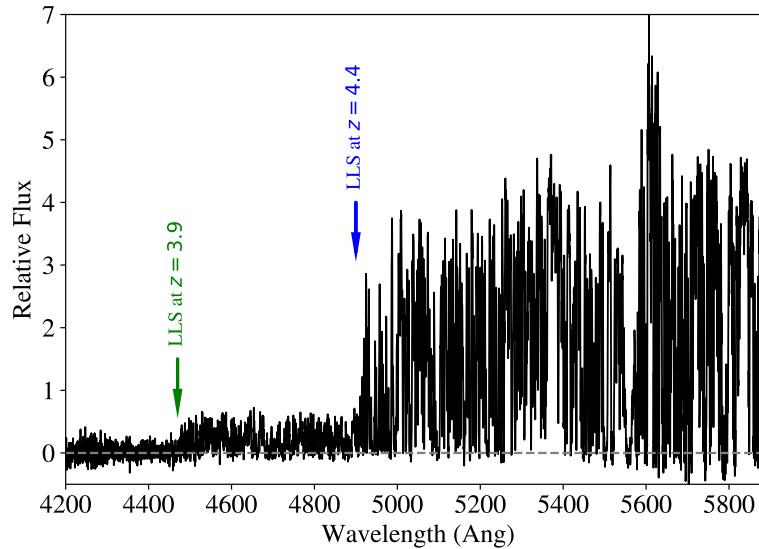


Fig. 32 A portion of the quasar spectrum of J0529–3526 recorded by the XQ-100 survey López et al (2016). One observes a sharp break in the flux at $\lambda \approx 4900\text{\AA}$ corresponding the Lyman limit opacity of an HI absorber at $z \approx 4.37$. The flux decrement corresponds to an optical depth of $\tau_{\text{LL}} \approx 1$. Further to the blue ($\lambda \approx 4470\text{\AA}$), one observes a second and complete break beyond which the quasar flux is completely absorbed. This second LLS at $z \approx 3.9$ has $\tau_{\text{LL}} > 3$.

and (3) the parent population density function, the probability that an LLS is observed in the interval $(t, t + \Delta t)$.

$$\lambda(t) = \frac{\text{Prob}(t < x < t + \Delta t | x > t)}{\Delta t} \text{ as } \Delta t \rightarrow 0 \quad (141)$$

Ansatz that the parent population is constant with t , i.e. $\lambda(t) = \lambda_0$. This is equivalent to a Poisson distribution for the number of LLS as a function of t . With $\lambda(t) = \lambda_0$, our survival function (Equation 140) evaluates trivially to $S(t) = \exp(-\lambda_0 t)$. This is akin to the decay of a particle with decay time ($\tau = 1/\lambda$) and the probability that the particle will survive for a time t . One may also derive $f(t) = \ell \exp(\lambda_0 t)$ and use (simple) statistical methods from biomathematics to derive the maximum likelihood for λ_0 from the observed distribution of LLS. We relate λ to the incidence of LLS at a given redshift, $\ell(z)_{\text{LLS}}$. Figure 34 shows the first estimates of this quantity from Tytler (1982).

The lower curve in Figure 34, defined there with notation $R(z)$, is the sensitivity function. In modern literature, it is notated with $g(z)$ where $g(z) dz$ is the number of sightlines that one may analyze for absorption in the interval $z, z + dz$. In essence, $g(z)$ expresses the ‘volume’ of any absorption line survey. In the Tytler (1982) survey (Figure 34), $R(z = 3) \approx 10$ indicates that the survey included 10 quasars whose

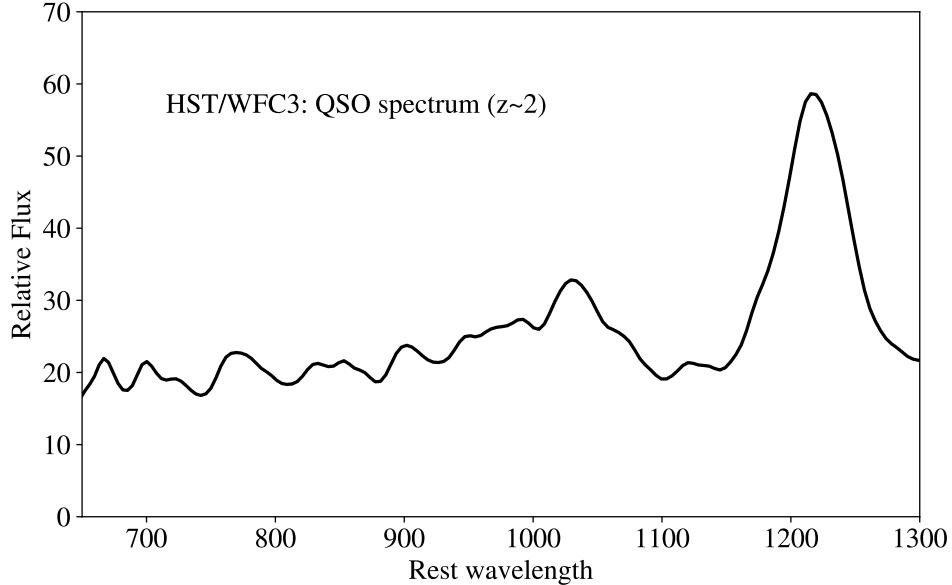


Fig. 33 Spectrum of a $z \approx 2$ quasar plotted in its rest-frame taken from the *HST/WFC3* survey of O’Meara et al (2013). This source shows a sustained and largely unattenuated flux blueward of its Lyman limit which enables the search for Lyman limit absorption from intervening gas (none is observed here).

spectra permitted the search for an LLS at $z = 3$. Again, in an LLS survey, the search along a given sightline runs from z_{em} to the maximum of $(z_{\text{min}}, z_{\text{LLS}})$ where z_{min} is set by the spectral coverage and z_{LLS} is the redshift of the first LLS identified along the sightline (if any).

It is also common practice to ignore systems close to the quasar by beginning the search several thousand km/s blueward of z_{em} . This accounts for quasar redshift error and minimizes any bias related to the ‘proximity’ of the quasar (e.g. an elevated incidence of LLS due to galaxy-galaxy clustering). An example of the selection function an LLS survey using the SDSS dataset (Prochaska et al, 2010) is shown in Figure 35. The shape of $g(z)$ is determined primarily by the quasar population recovered by the survey.

5.4 Incidence of LLS: $\ell_{\text{LLS}}(z)$

Another approach to estimating $\ell_{\text{LLS}}(z)$, motivated by survival analysis, is through a binned evaluation of the number of systems discovered relative to the integrated sensitivity function (i.e. survey path). Consider, $\ell_{\text{LLS}}(z)$ in a redshift interval $[z_1, z_2]$. The estimator is

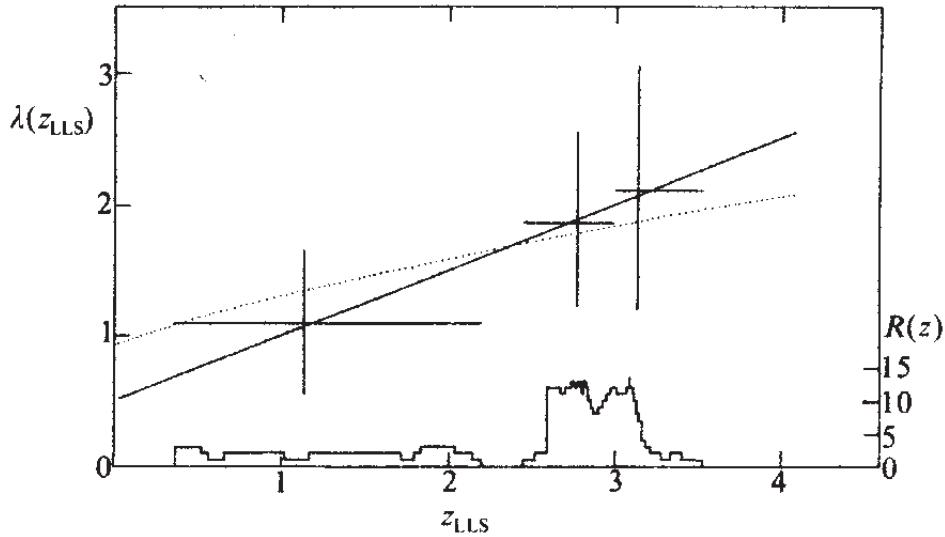


Fig. 34 Binned evaluations of the incidence of LLS per unit redshift (here expressed as $\lambda(z_{\text{LLS}})$) from Tytler (1982). The histogram describes the sensitivity function of their survey, i.e. the number of quasars searched for LLS at a given redshift.

$$\ell_{\text{LLS}}(z) = \frac{\text{Number of systems detected in } [z_1, z_2] (\mathcal{N})}{\int_{z_1}^{z_2} g(z) dz} \quad (142)$$

One estimates uncertainty using the variance from Poisson statistics: $\sigma^2(\mathcal{N})$. Figure 36 shows an example of results from $z \sim 4$ for LLS with optical depths $\tau_{\text{LL}} \geq 2$.

We may model $\ell_{\text{LLS}}(z)$ independent of binning, with the Maximum Likelihood technique described in section 4.3 for $f(N_{\text{HI}})$ (but without the N_{HI} dependence). For the same arguments made on the Ly α Forest, we may expect the redshift evolution to scale as $(1+z)^\gamma$. Ignoring the N_{HI} dependence in Equation 82, the log-Likelihood probability for a given sightline is given by:

$$\ln \mathcal{L}_j = \sum_{i=1}^M -\mu_i + \sum_{k=1}^p \ln \mu_k \quad (143)$$

$$= \sum_i^M -\ell(z_i) dz + \sum_k^p \ln \ell(z_k) \quad (144)$$

This expression is for the j th quasar and we have related the expected number of LLS in a cell with ‘volume’ dz to be $\mu_i = \ell(z_i) \delta z$. We also recognize that for an LLS survey $p = 0$ or 1 because we are limited to detecting one LLS per sightline.

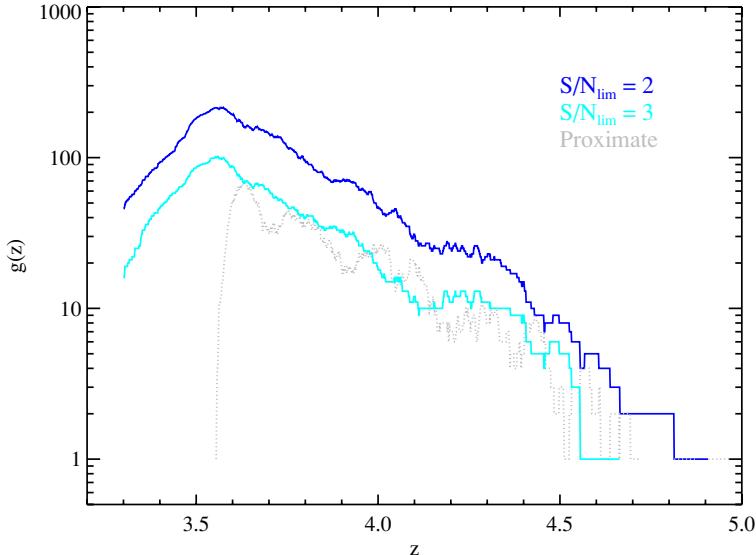


Fig. 35 Sensitivity function of Prochaska et al (2010), cut by S/N. The increase in $g(z)$ at low redshifts is due to their avoiding $z < 3.5$ quasars which are biased in SDSS to exhibit LLS (see section 5.5). The monotonic decrease in $g(z)$ beyond $z \approx 3.5$ is due to the decreasing number of quasars with high quality spectra in the survey.

Letting $dz \rightarrow 0$,

$$\ln \mathcal{L}_j = \int_{z_{\text{start}}}^{z_{\text{end}}} \ell(z) dz + \sum_k^p \ln \ell(z_k), \quad (145)$$

The full log-likelihood $\ln \mathcal{L}$ is the sum over all quasars. Maximizing $\ln \mathcal{L}$ yields the best parameters for a given model. We may then assess the goodness of the model by comparing the observed cumulative distribution of LLS redshifts against that predicted by the model after integrating over $g(z)dz$. This is frequently assessed with a one-sided Kolmogorov-Smirnov (KS) test. This is shown in Figure 36) with $P_{\text{KS}} \sim 1$ indicating a fully acceptable model.

Results on $\ell_{\text{LLS}}(z)$ across a wide redshift interval are shown in Figure 37, again for LLS with $\tau_{\text{LL}} \geq 2$ (Ribaudo et al, 2011). The data have been modeled with a power-law of the form

$$\ell(z) = \ell_* \left[\frac{1+z}{1+z_*} \right]^\gamma \quad (146)$$

and the authors report $z_* = 3.23$, $\ell_* = 1.62$, $\gamma = 1.83 \pm 0.21$.

Consider the distance Δr that one travels on average to intersect one LLS at $z = 3.5$. The redshift path is approximately $\Delta z \approx 1/\ell(z = 3.5) \approx 0.56$. From cosmology,

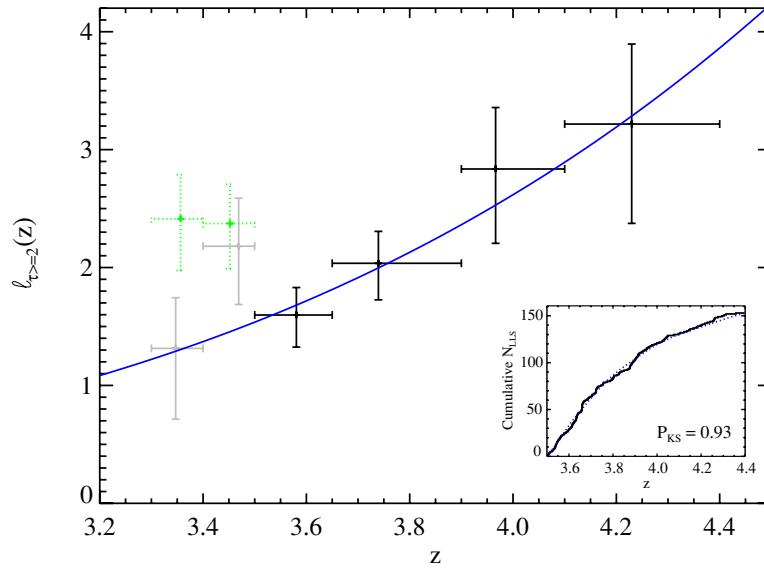


Fig. 36 Incidence of LLS with $\tau_{\text{LL}} \geq 2$ per unit redshift $\ell_{\text{LLS}}(z)$ from $z \approx 3.3 - 4.4$ derived from quasar spectra in the SDSS DR7 Prochaska et al (2010). The black binned evaluations show an increasing incidence with redshift. The inset compares the observed, cumulative distribution of LLS redshifts against the model prediction. The green dotted points show $\ell(X)$ for the full set of SDSS quasars observed in the survey. Because of their color selection (see section 5.5), the quasars with $z_{\text{em}} < 3.5$

the physical distance is

$$\Delta r \approx \frac{c\Delta z}{H(z)(1+z)} \quad (147)$$

and we estimate $\Delta r \approx 100$ Mpc (see also the *Optically_Thick* Notebook). This exceeds the mean separation between faint quasars by a factor of 3 – 10 Faucher-Giguère et al (2009). Therefore, we expect each volume in the universe sees the radiation from multiple sources, although we caution that the opacity to ionizing photons may be dominated by gas with $\tau_{\text{LL}} < 2$. We discuss a more accurate measurement of the mean free path below.

5.5 Survey Subtleties

Thus far we have ignored any systematic effects in HI absorption analysis related to the selection of background sources. For the Ly α forest, it is possible that these are

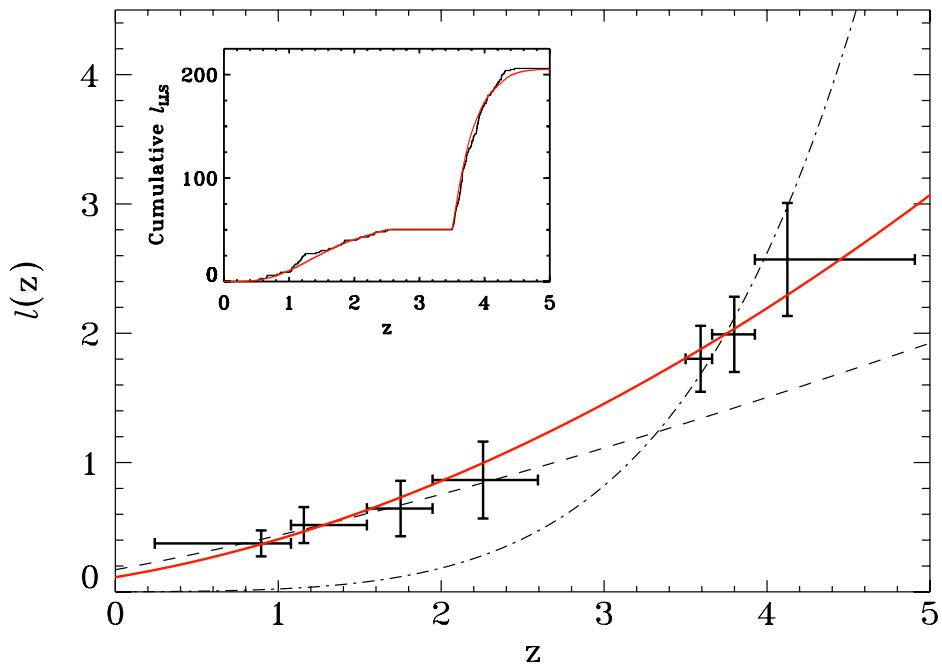


Fig. 37 Incidence of LLS with $\tau_{\text{LL}} \geq 2$ per unit redshift $\ell_{\text{LLS}}(z)$ from $z \approx 0 - 5$ combining surveys with the *HST* and SDSS (Ribaudo et al, 2011) derived from quasar spectra in the SDSS DR7 Prochaska et al (2010). The red curve, derived with Maximum likelihood techniques, is a $(1+z)^\gamma$ power-law with $\gamma = 1.83$ which describes the data well across cosmic time.

largely negligible. But for stronger HI absorption, the gas has substantial effect on the observed source flux.

Indeed, LLS attenuate the color of the background source and may bias their selection. For a quasar at $z = 3$, the u -band flux is greatly diminished by the continuum opacity of an LLS. Projects which select by optical (rest-frame UV) color are then sensitive to the presence of an intervening LLS. For quasars, this is particularly an issue at $z \sim 3$ when their colors lie close to the stellar locus (see Figure 38). In general, the two are indistinguishable using only optical photometry. But if the intrinsic color is highly altered, e.g. the u -band flux of a quasar is absorbed by an intervening LLS, the source separates from the stellar locus and may be targeted. Indeed, the SDSS quasars at $z \sim 3$ exhibit a higher incidence of LLS than the true parent population (see Prochaska et al (2010) and Figure 35).

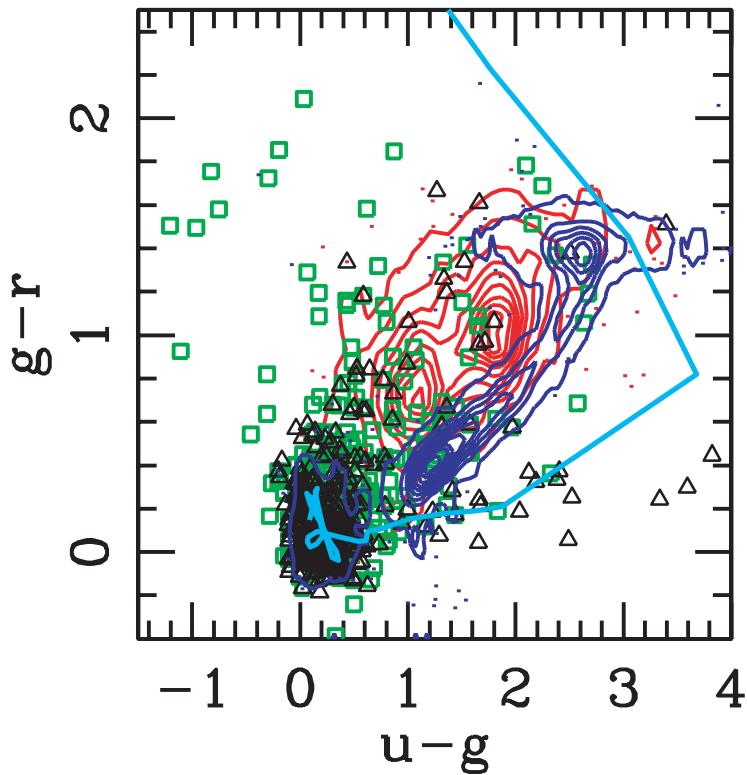


Fig. 38 Color-color plot of point sources in SDSS photometry (contours), primarily Galactic stars and white dwarfs. The black triangles are spectroscopically confirmed Type 1 quasars and the cyan curve is the color-color track for quasars from $z=0$ to 7 . In this parameter space, quasars only separate themselves from stars when they have a red $u-g$ color. This occurs for $z = 3$ quasars when an intervening LLS absorbs the u -band flux. Taken from Richards et al (2006).

In short, when one achieves high statistical precision from today's large datasets, systematic biases like these become relevant and must be carefully considered.

5.6 Absorption path $X(z)$

Bahcall and Peebles (1969) introduced a pathlength definition as a means to test cosmological models with the measured incidence of absorption-line systems. The basic concept is to define $X(z)$ such that the incidence of absorbers per dX is *constant* provided the product of the following two quantities is constant: (a) the comoving number density of absorbers $n_c(z)$; and (b) the physical cross-section $A_p(z)$. Under the (brazen) assumption of a non-evolving population, one can then solve for the cosmological parameters that yield a constant $n_c A$ product.

We nearly derived $X(z)$ in our discussion of redshift evolution for the line density in the Ly α forest (section 4.5). Recall, we noted that the incidence of absorption per proper distance is

$$\frac{d\mathcal{N}}{dr_p} = n_p(z)A_p(z) \quad (148)$$

with n_p the proper density of absorbers. Expressed in terms of dz , we have

$$\frac{d\mathcal{N}}{dz} = \ell(X) = n_p(z)A_p(z) \frac{c}{H(z)(1+z)} \quad (149)$$

Now introduce dX such that,

$$\ell(X) = C n_c(z) A_p(z) , \quad (150)$$

with C a constant. In an expanding universe, $n_c(z) = n_p(z)/(1+z)^3$. And, if we wish dX to be a dimensionless quantity (like redshift), we choose

$$dX = \frac{H_0}{H(z)}(1+z)^2 dz \quad (151)$$

to yield

$$\ell(X) = \frac{c}{H_0} n_c(z) A_p(z) \quad (152)$$

As desired, this quantity is static with an expanding universe provided the comoving number density of sources is constant and their physical size does not evolve.

In Figure 39, we show the dX/dz curves versus redshift for a few cosmologies. Two trends are apparent: (1) there is significant evolution in dX/dz with redshift and (2) dX/dz is sensitive to the cosmology (but not H_0). Therefore, a non-evolving population of absorbers (if any such population exists!) offers significant sensitivity to cosmology. Alternatively, we may adopt our concordance Λ CDM cosmology (e.g. Planck2015) and interpret evolution in $\ell(X)$ as intrinsic evolution in the gas.

We can derive $\ell_{LLS}(X)$ for the LLS by repeating the Maximum likelihood analysis after replacing dz with dX . All other aspects of the analysis are identical. Figure 40 shows the results across cosmic time. We identify a steep decline with decreasing redshift that demands an evolving population. Either the comoving number density of optically thick sources greatly declines (n_c) and/or their physical size (A_p) is decreasing.

Consider a toy model of dark matter halos within hierarchical cosmology (Fumagalli et al, 2013). This model yields a prediction for $\ell(X)$ with redshift according to assumptions made on the covering fraction f_c of optically thick gas in each halo.

$$\ell(X) = \frac{4c}{H_0} \int_{\log M_{\text{low}}}^{\log M_{\text{up}}} R_{\text{vir}}^2(M_{\text{vir}}, z) f_c(M_{\text{vir}}, z) \frac{dn_c}{d \log M_{\text{vir}}} d \log M_{\text{vir}} \quad (153)$$

Here $M_{\text{low}}, M_{\text{up}}$ define the mass range of halos contributing to LLS and we recognize $A_p = f_c \pi R_{\text{vir}}^2$ with R_{vir} the virial radius. The last term in Equation 153 expresses the comoving number density. Figure 41 compares this toy model against the observations for several assumptions on f_c with each tuned to match at $z \approx 3$ (Fumagalli et al, 2013). The red dotted line in the Figure assumes constant f_c with redshift. Because of the growth of structure with decreasing redshift in our cosmology, this

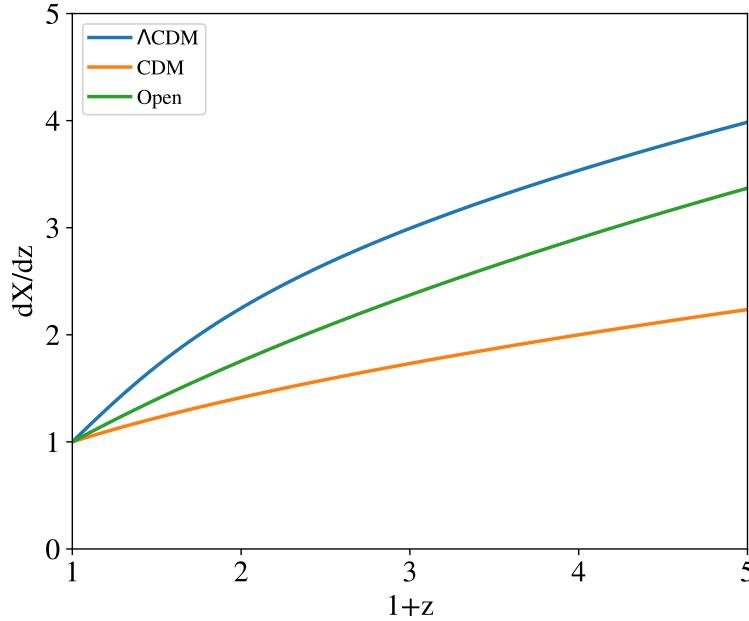


Fig. 39 dX/dz curves for a range of cosmologies. The significant differences imply sensitivity to the underlying cosmological parameters (Bahcall and Peebles, 1969).

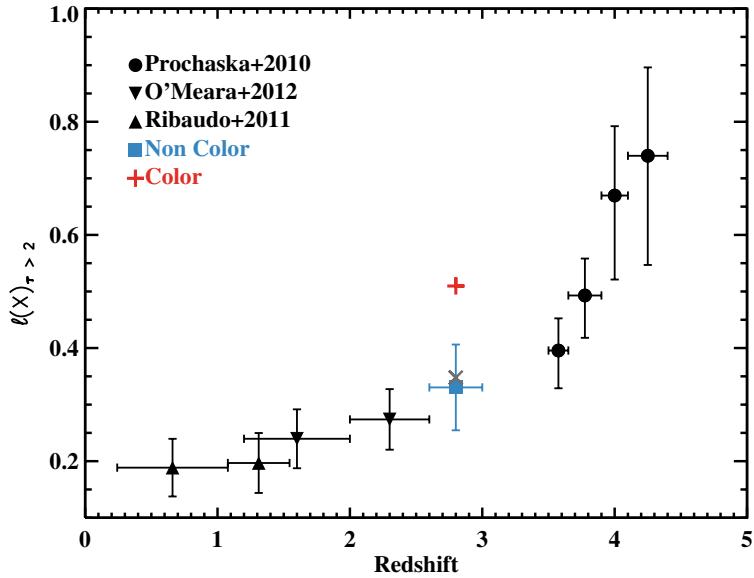


Fig. 40 Redshift evolution of the incidence of LLS $\ell_{\text{LLS}}(X)$ with $\tau_{\text{LL}} \geq 2$ per absorption path length dX . The red and blue points distinguish between quasar samples selected on the basis of color or not (respectively); focus on the latter. If the structures giving rise to LLS had a static comoving number density and physical size, $\ell(X)$ would be constant in time (provided we have adopted the correct cosmology). The steep increase in $\ell(X)$ with z indicates substantial intrinsic evolution in the structures hosting optically thick gas. Taken from Fumagalli et al (2013).

model predicts a steep *increase* in $\ell(X)$ towards the present-day that strongly violates the observed evolution.

The blue and green lines are simple “wind” and gas accretion scenarios which are tuned to describe the observations at $z < 3$. These also fail at high- z and one infers that the IGM, and not dark matter halos, contributes greatly to optically thick gas at $z > 3.5$.

5.7 Mean Free Path $\lambda_{\text{mfp}}^{912}$

Optically thick gas attenuates the flux of ionizing photons throughout the Universe. The LLS act, in essence, as brick walls absorbing all photons with $h\nu > 1$ Ryd. However, gas with $N_{\text{HI}} \lesssim 10^{17} \text{ cm}^{-2}$ will also attenuate the ionizing flux and, if sufficiently common, can dominate the absorption. The integrated effect of all gas yields an HI mean free path to ionizing photons.

Cosmologically, the mean free path is a fundamental input to calculations of the extragalactic UV background (EUVB). It relates the emissivity of sources to the mean intensity (Haardt and Madau, 1996; Faucher-Giguère et al, 2009):

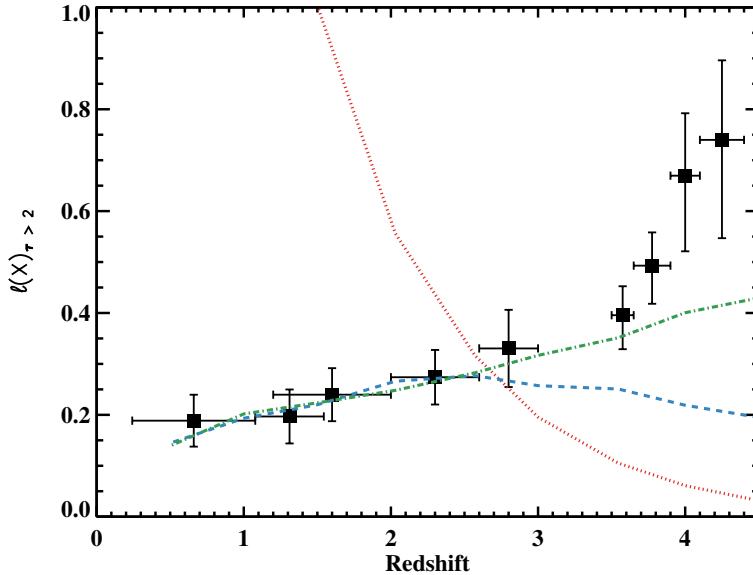


Fig. 41 Comparison of the $\ell_{\text{LLS}}(X)$ data from Figure 40 with toy models for the evolution in $\ell(X)$ assuming LLS arise primarily in dark matter halos. The curves correspond to fixed covering fraction of optically thick gas (red), a wind model that ejects the gas (blue) and an accretion scenario that also reduces the covering fraction with decreasing redshift (green). The latter two models offer a good description of the observations for $z < 3$, but fail at higher redshifts. These results argue for an increasing contribution to the Lyman limit opacity from structures in the IGM. Taken from Fumagalli et al (2013).

$$J_\nu(z) \approx \frac{1}{4\pi} \lambda_{\text{mfp}}^{912} \epsilon_\nu(z) \quad (154)$$

The mean free path is also relevant to calculations of the topology of HI reionization at $z > 5$, altering the leakage of ionizing sources into the IGM. Currently, $\lambda_{\text{mfp}}^{912}$ cannot be predicted from first principles and one must constrain its value (and evolution) with observation.

Let us define $\lambda_{\text{mfp}}^{912}(z_{\text{em}})$ as the physical distance that a packet of photons travel from a source at $z = z_{\text{em}}$ before suffering an e^{-1} attenuation. In contrast to other physical quantities, this definition is directly coupled to the source redshift. If we define an effective optical depth for Lyman continuum opacity $\tau_{\text{eff}}^{\text{LL}}$, then $\lambda_{\text{mfp}}^{912}$ is the distance from the source where $\tau_{\text{eff}}^{\text{LL}} = 1$. There is, however, an important and subtle concept associated with defining and measuring $\lambda_{\text{mfp}}^{912}$ due to the expansion of the universe. Consider a packet of photons with energy of 1 Ryd emitted by the source. These will travel only an infinitesimal distance⁸ before redshifting beyond 1 Ryd to become insensitive to the HI continuum opacity. This means that the packet of

⁸ In reality, peculiar motions of the gas including Doppler broadening imply a greater than infinitesimal distance.

photons which suffers an e^{-1} attenuation must have rest-frame energy $h\nu > 1$ Ryd. Analysis of $\lambda_{\text{mfp}}^{912}$ therefore requires a proper treatment of cosmic expansion and the frequency dependence of the continuum opacity.

The traditional approach to estimate $\lambda_{\text{mfp}}^{912}$ was to calculate $\tau_{\text{eff}}^{\text{LL}}$ from our $f(N_{\text{HI}})$ distribution and relate $\tau_{\text{eff}}^{\text{LL}} = 1$ to $\lambda_{\text{mfp}}^{912}$. Here, $\tau_{\text{eff}}^{\text{LL}}$ is the effective optical depth from Lyman continuum opacity that a photon of $\nu \geq \nu_{912}$ experiences, $F_{\text{obs}} = F \exp(-\tau_{\text{eff}}^{\text{LL}})$. In contrast to the effective opacity from Lyman series lines (e.g. $\tau_{\text{eff},\alpha}$), $\tau_{\text{eff}}^{\text{LL}}$ is an integral (cumulative) quantity. That is, it depends on the IGM along the entire path traveled until the photon with frequency ν redshifts to ν_{912} at

$$z_{912} \equiv (1 + z_{\text{em}}) \frac{\nu_{912}}{\nu} - 1 \quad (155)$$

with $\nu_{912} = 1$ Ryd/h.

The derivation from $f(N_{\text{HI}})$ is as follows. The fractional attenuation from Lyman continuum opacity for gas with N_{HI} is

$$1 - \exp[-N_{\text{HI}} \sigma_{\text{photo}}(\nu)] \quad (156)$$

i.e. a transparent source gives null attenuation and a ‘wall’ gives complete attenuation. For a population of absorbers, we weight by $f(N_{\text{HI}})$. Lastly, we must account for the frequency dependence of σ_{photo} (using Equation 155) to recover

$$\tau_{\text{eff}}^{\text{LL}}(z_{912}, z_{\text{em}}) = \int_{z_{912}}^{z_{\text{em}}} \int_0^{\infty} f(N_{\text{HI}}, z') \{1 - \exp[-N_{\text{HI}} \sigma_{\text{ph}}(z')]\} dN_{\text{HI}} dz' \quad (157)$$

For the $f(N_{\text{HI}})$ model of Prochaska et al (2014), we calculate $\tau_{\text{eff}}^{\text{LL}}$ as shown in Figure 42. Of course, $\tau_{\text{eff}}^{\text{LL}}$ rises with decreasing redshift as the photons travel to ever greater distances from the source. We see that $\tau_{\text{eff}}^{\text{LL}} = 1$ at $z \approx 3.2$ and the proper distance from $z = 3.5$ gives $\lambda_{\text{mfp}}^{912} \approx 60$ Mpc.

This technique of integrating $f(N_{\text{HI}})$ is indirect, i.e. we have predicted the integrated Lyman continuum from analysis of the HI Lyman series opacity of the Ly α forest. But $f(N_{\text{HI}})$ at $N_{\text{HI}} \approx 10^{17} \text{ cm}^{-2}$ is highly uncertain because the Lyman series lines lie on the saturated portion of the COG. Furthermore, line-blending poses a significant problem and surveys of $\tau \approx 1$ LLS are difficult to perform. Indeed, the $f(N_{\text{HI}})$ function of Prochaska et al (2014) used an estimation of $\lambda_{\text{mfp}}^{912}$ as a constraint at $N_{\text{HI}} \lesssim 10^{17} \text{ cm}^{-2}$.

Prochaska et al (2009) proposed a new method to directly assess the average Lyman continuum opacity $\tau_{\text{eff}}^{\text{LL}}$ and thereby $\lambda_{\text{mfp}}^{912}$. A similar concept was introduced in section 4.7 where $\tau_{\text{eff},\alpha}$ was estimated from a composite spectrum. By averaging the IGM absorption along many sightlines, the composite spectrum (without weighting) provides an optimal means to measure the distance traveled for an e^{-1} attenuation. The concept is illustrated in Figure 43.

In practice, we model the IGM attenuation with an effective opacity κ_{LL} for Lyman continuum absorption

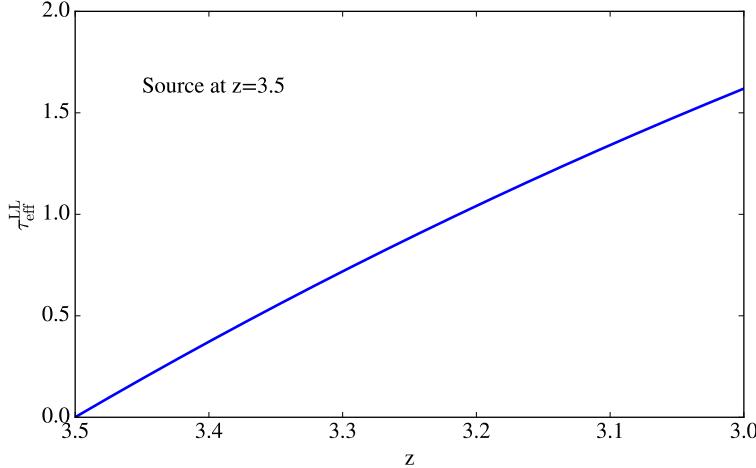


Fig. 42 Effective opacity as a function of $z < z_{\text{em}}$ with $z_{\text{em}} = 3.5$. In contrast to $\tau_{\text{eff}}^{\text{Ly}\alpha}$ (or the effective opacity of any Lyman series line), $\tau_{\text{eff}}^{\text{LL}}$ is a cumulative quantity and must increase monotonically with decreasing redshift from the source.

$$\kappa_{\text{LL}}(z, v) = \tilde{\kappa}_{912}(z) \left(\frac{v}{v_{912}} \right)^{-2.75}. \quad (158)$$

The frequency term captures the energy dependence of the photoionization cross-section and κ_{912} is a normalization which captures IGM evolution with redshift. Integrating this opacity over a physical distance and allowing for redshift of the photons

$$\tau_{\text{eff}}^{\text{LL}} = \int_0^r \kappa_{\text{LL}}(r') dr' \quad (159)$$

$$= \int_{z_{912}}^{z_{\text{em}}} \kappa_{\text{LL}}(z', v) \frac{dr}{dz} dz' \quad (160)$$

As demonstrated above, at $z > 2$ we have $dr/dz \propto (1+z)^\beta$ with $\beta \approx -2.5$. Our model reduces to a simple integral of $(1+z)$ to some power and one additional parameter (κ_{912}). This provides great statistical power (i.e. a one parameter fit to many pixels). The primary uncertainties are our estimate for the quasar SED and sample variance (estimated from a bootstrap approach).

Model fits to composite quasar spectra at $z \sim 4$ are shown in Figure 44 from Prochaska et al (2009). The mean free path measured from the direct evaluation is larger than predicted from older $f(N_{\text{HI}})$ models. In turn, it implies less attenuation of ionizing sources and therefore fewer ionizing sources required to generate the intensity of the EUVB.

Similar analysis has been performed at a range of redshifts (O'Meara et al, 2013; Fumagalli et al, 2013; Worseck et al, 2014). Each composite yields an estimate of $\lambda_{\text{mfp}}^{912}$ by estimating the distance where the flux is attenuated by e^{-1} . These are presented in Figure 45. Assuming a $(1+z)^\gamma$ power-law evolution in $\lambda_{\text{mfp}}^{912}$, the authors report

$$\lambda_{\text{mfp}}(z) = (37 \pm 2 h_70^{-1} \text{Mpc}) \left[\frac{1+z}{5} \right]^{-5.4 \pm 0.4} \quad (161)$$

As with the LLS, this is far steeper than cosmological expansion alone would predict. It requires the disappearance of optically thick gas with decreasing redshift. This likely results from the continued ionization of the universe as sources (quasars) rise in number and/or the accretion of optically thick gas onto galaxies.

These new results on $\lambda_{\text{mfp}}^{912}$ also impact estimations for the escape fraction of ionizing photons from galaxies f_{esc} . The observational experiment is to directly observe the flux of a galaxy (e.g. at $z=3$) at wavelengths just blueward of its Lyman limit. For this, one employs a narrow-band filter (e.g. Nestor et al, 2013) or integrates a spectrum near the the Lyman limit: $\langle f_\nu \rangle_{\text{LL,obs}}$. The escape fraction is then defined as the flux of radiation that escapes versus that which the galaxy produced (from O and B stars):

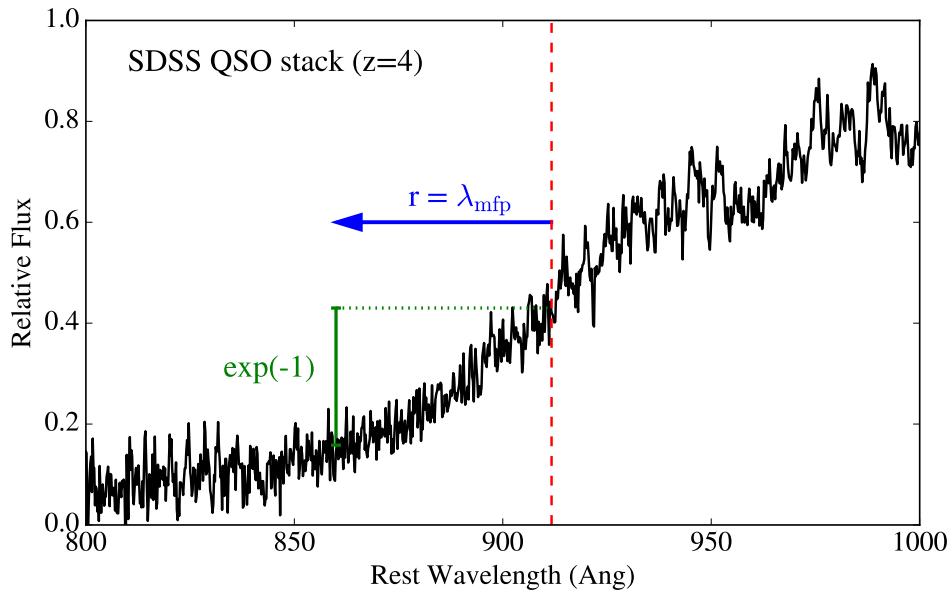


Fig. 43 Composite spectrum of ~ 150 quasars at $z_{\text{em}} \approx 4$ (black). The declining flux at $\lambda < 911.7 \text{\AA}$ (red dashed line) is due to the cumulative effective opacity of Lyman limit absorption $\tau_{\text{eff}}^{\text{LL}}$. By estimating the redshift where the flux has suffered an e^{-1} attenuation, one directly measures $\lambda_{\text{mfp}}^{912}$.

$$f_{\text{esc}} = \frac{\langle f_v \rangle_{\text{LL,obs}}}{\langle f_v \rangle_{\text{LL,intrinsic}}} \quad (162)$$

One is interested here in the flux attenuation by gas in the galaxy, but $\langle f_v \rangle_{\text{LL,obs}}$ is also attenuated by the IGM.

As an example, a medium-band filter ($\text{FWHM} \approx 100\text{\AA}$) placed just blueward of a $z = 3.5$ galaxy's Lyman limit covers $z = 3.4$ to 3.5 at the Lyman limit. We estimate the IGM effective opacity to be $\langle \tau_{\text{eff}}^{\text{LL}} \rangle = 0.25$ which gives a 30% correction to f_{esc} that increases steeply with redshift.

If we extrapolate the evolution in $\lambda_{\text{mfp}}^{912}$ to $z < 2$, we find that $\lambda_{\text{mfp}}^{912}$ exceeds the Horizon at $z \approx 1.6$. This is referred to as the “breakthrough” redshift where photons are no longer (significantly) attenuated by the IGM. Beyond this epoch, every ionizing source can see every other.

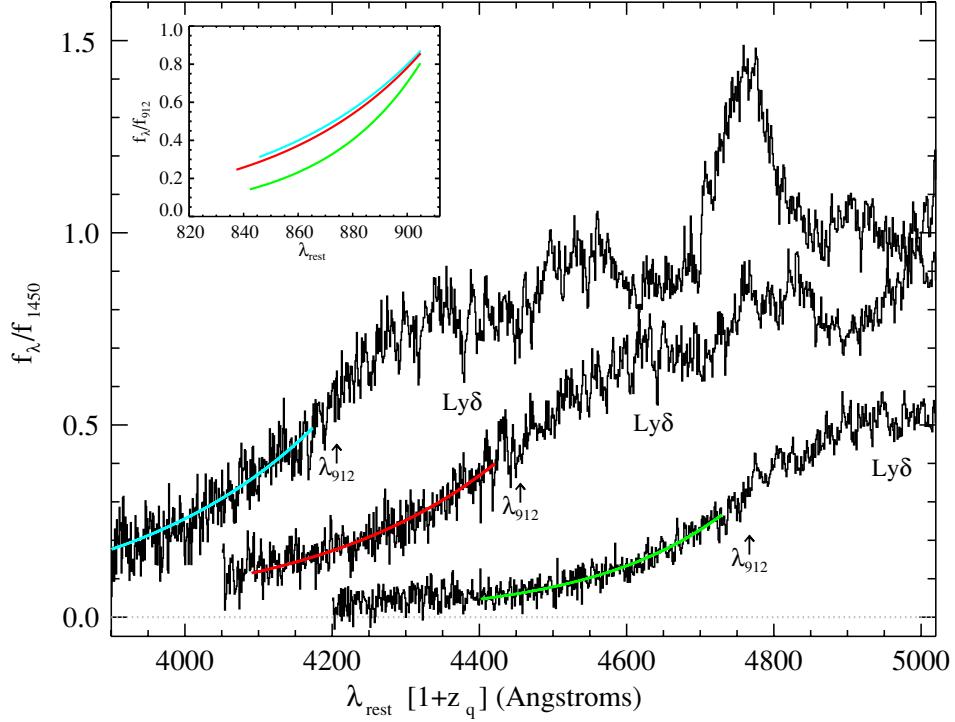


Fig. 44 Composite spectra for a series of $z_{\text{em}} \approx 3.8 - 4.5$ quasars from the SDSS (black curves). Overplotted on the data are models of the IGM attenuation from HI continuum opacity. The e^{-1} attenuation measured from these models gives mean free path $\lambda_{\text{mfp}}^{912}$ measurements. Taken from Prochaska et al (2009)

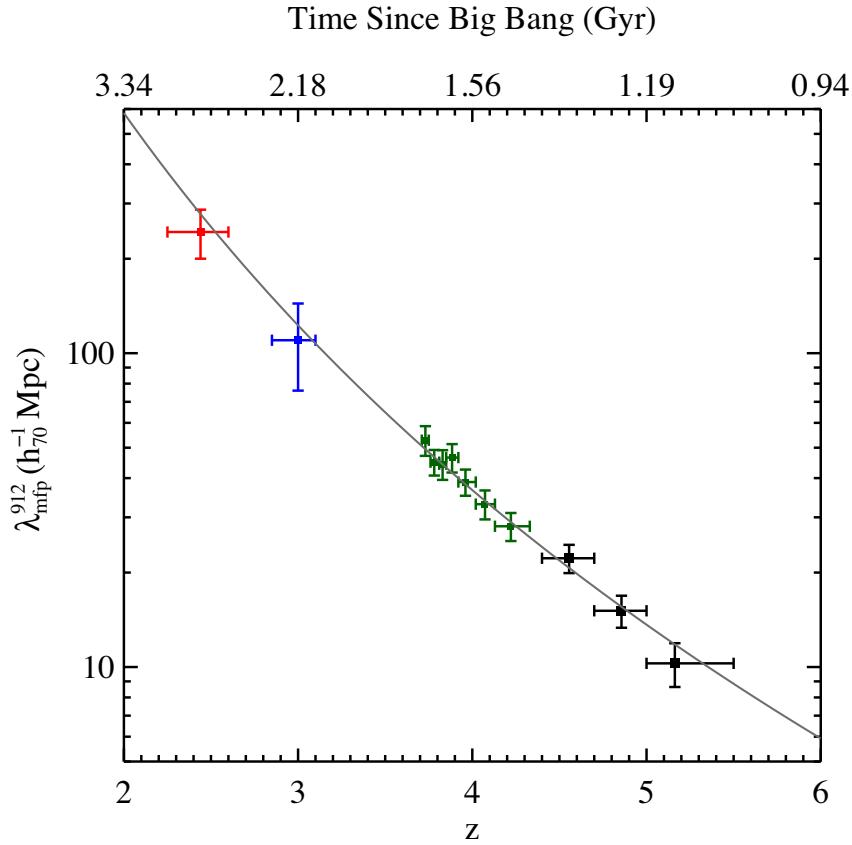


Fig. 45 Series of $\lambda_{\text{mfp}}^{912}$ measurements at a series of source redshifts from analysis of composite spectra. The overplotted model is a $(1+z)^\gamma$ fit with $\gamma \approx -5.4$. This very steep evolution in $\lambda_{\text{mfp}}^{912}$ implies ongoing ionization of gas in the IGM down to $z < 3$. Taken from Worseck et al (2014).

5.8 Connecting LLS to Theory

We conclude our discussion of LLS with a few additional comments on the origin of optically thick gas in a cosmological context. Early cosmological simulations woefully underpredicted the incidence of LLS (Gardner et al, 2001). In hindsight, this was not surprising; the requirements to model LLS on simulations are steep. LLS are related to non-linear structure formation and are sensitive to the radiative transfer of ionizing photons. This is a terrific challenge in terms of resolution (one needs kpc-scale sensitivity over hundreds of Mpc) and as regards the physics one can implement in cosmological, hydrodynamic simulations.

More recently, theory has connected LLS to an emerging paradigm for accreting cold ($T \sim 10^4$) gas into dark matter halos, a.k.a. cold streams or cold flows (e.g.

Dekel and Birnboim, 2006; Kereš et al, 2009). Such gas may be optically thick (or partially optically thick) and a series of papers have associated at least a fraction of LLS with these streams (e.g. Faucher-Giguère and Kereš, 2011; Fumagalli et al, 2011; van de Voort and Schaye, 2012). An illustration of the results is shown in Figure 46. Here one finds extended, optically thick gas throughout the underlying, dark matter halo. One estimates that a modest ($\approx 10 - 20\%$) fraction of the halo is covered by optically thick gas which is metal-poor but not primordial. This leads to a significant number of LLS, but it remains difficult to reconcile the incidence of LLS with dark matter halos alone (see the $\ell(X)$ measurements above). Again, at $z > 3$ one speculates that the dominant contribution to LLS is from gas in large-scale structure (e.g. filaments). Ultimately, this is a fundamental benchmark for future cosmological simulations with a full treatment of radiative transfer.

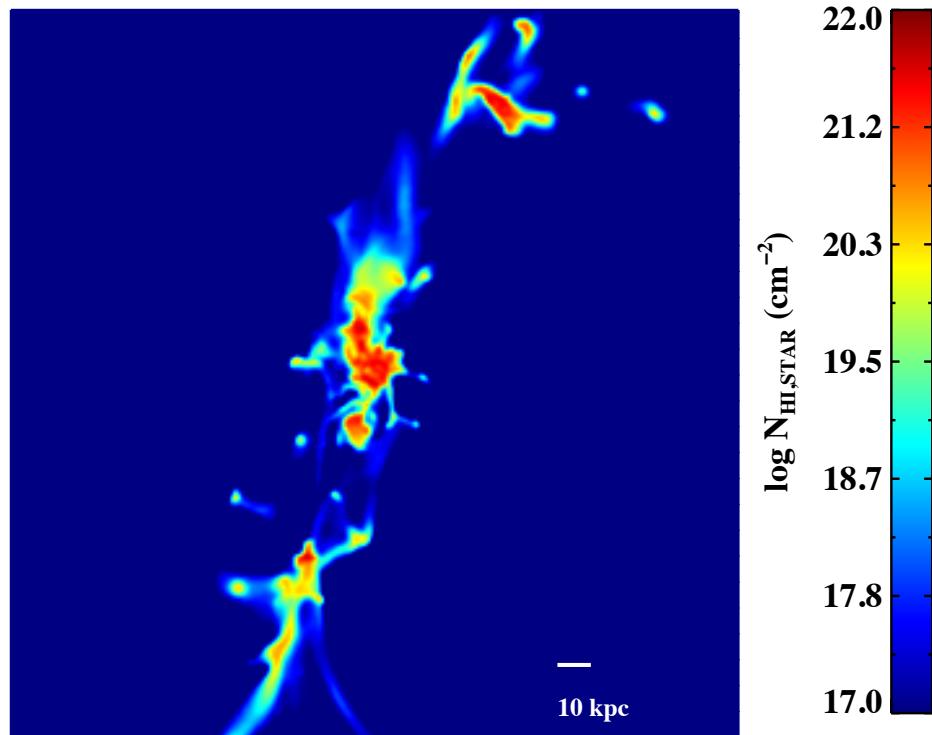


Fig. 46 Visualization of the HI column density of streams flowing into a dark matter halo at $z \approx 3$. At the center of the image is a star-forming galaxy and its ISM dominates the HI gas. Beyond ≈ 10 kpc from the galaxy, the gas is primarily related to streams and dwarf galaxies within those streams. These structures are predicted to dominate the cross-section of optically thick gas within dark matter halos. Taken from Fumagalli et al (2011).

5.9 Damped Ly α Systems

In the school, I presented a short lecture on the damped Ly α systems (DLAs), HI gas with $N_{\text{HI}} \geq 2 \times 10^{20} \text{ cm}^{-2}$. Obviously, these systems are also optically thick at the Lyman limit and are therefore LLS with $\tau_{\text{eff}}^{\text{LL}} \gg 1$. Readers interested in DLAs are referred to the review article by Wolfe et al (2005) (and my DLA lecture). I include only a few key observational results here.

The frequency distribution $f(N_{\text{HI}})$ of the DLAs has been well estimated from the large datasets of SDSS and BOSS. Using the SDSS DR5, Prochaska et al (2005) established that $f(N_{\text{HI}})$ in the DLA regime follows a single power-law with $\alpha \approx -2$ until an observed break at $N_{\text{HI}} \approx 10^{21.5} \text{ cm}^{-2}$ (Figure 47). This break may be naturally understood in a simple model of DLAs as a population of randomly inclined exponential disks (Fall and Pei, 1993; Wolfe et al, 1995). Assuming a face-on column density distribution with radius r

$$N_{\perp}(r) = N_{\perp,0} \exp(-r/r_d) \quad (163)$$

with $N_{\perp,0}$ the central column density and r_d a scale-length. For random, face-on disks (no inclination) the probability of intersection at r is proportional to r giving

$$f(N_{\perp}, X) = \frac{2\pi c r_d^2 n_c}{H_0} \left[\frac{\ln(N_{\perp,0}/N_{\perp})}{N_{\perp}} \right] \quad (164)$$

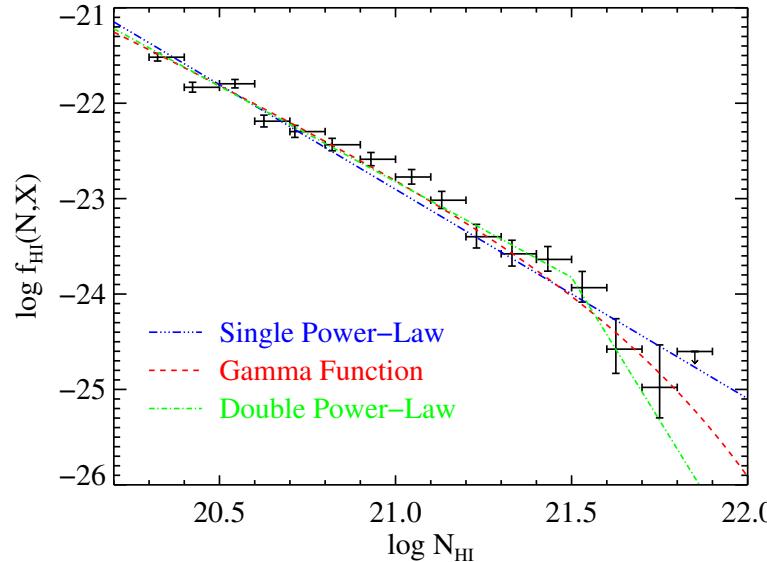


Fig. 47 Frequency distribution of N_{HI} for the damped Ly α systems at $z \approx 2.5$. These gas layers show a roughly power-law distribution with $\alpha \approx -2$ for lower N_{HI} values which breaks to a steeper function at $N_{\text{HI}} > 10^{21.5} \text{ cm}^{-2}$. Taken from Prochaska et al (2005).

with n_c is the assumed comoving number density of DLAs. Lastly, allow for random inclination

$$f(N, X) = \int_0^{\min(N, N_{\perp,0})} dN_{\perp} \left[\frac{N_{\perp}^2 f(N_{\perp})}{N^3} \right] \quad (165)$$

Evaluating

$$f(N, X) = \begin{cases} \frac{1}{N} \left[1 - 2 \ln \left(\frac{N}{N_{\perp,0}} \right) \right] & N \leq N_{\perp,0} \\ \frac{N_{\perp,0}^2}{N^3} & N \geq N_{\perp,0} \end{cases} \quad (166)$$

The first term scales as N^{-2} for $N \ll N_{\perp,0}$ i.e., we expect a broken power-law!

A larger sample of DLAs has been analyzed by Noterdaeme et al (2012) from BOSS. The report an approximately N_{HI}^{-3} decline at the highest column densities and the discovery of DLAs towards quasars with $N_{\text{HI}} > 10^{22} \text{ cm}^{-2}$.

Surprisingly, the $f(N_{\text{HI}})$ distribution for DLAs shows very weak evolution in its shape across cosmic time. This includes $z = 0$, i.e. 21 cm observations (Zwaan et al, 2005), but see Braun (2012). This implies a nearly non-evolving population of HI gas from $z = 2$ to 0.

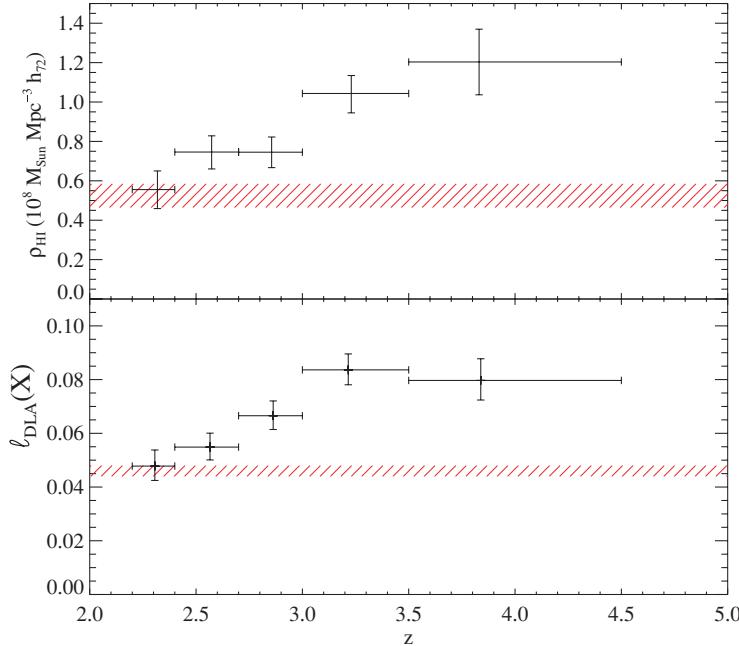


Fig. 48 Estimates for the incidence of DLA absorption (lower) and the total mass density in atomic gas contained within these systems ρ_{HI} (upper). The red bands express the same quantities measured at $z \sim 0$ from 21 cm observations (Zwaan et al, 2005). Taken from Prochaska and Wolfe (2009).

Measurements of $\ell_{\text{DLA}}(X)$ constrain the evolution of HI gas within (and around) galaxies across cosmic time (Figure 48). One observes a sharp evolution from $z = 2 - 4$ which requires a $2 \times$ decline in $n_c A_p$ in this ≈ 2 Gyr window. It remains an open question whether this evolution is associated with photoionization of the gas, its rapid accretion from the halo onto galaxies, and/or processes related to galaxy-galaxy mergers.

5.10 $f(N_{\text{HI}})$ Revisited

We may integrate the results described thus far in the Chapter to model $f(N_{\text{HI}})$ for the IGM (approximately $N_{\text{HI}} < 10^{15} \text{ cm}^{-2}$), the optically thick gas traced by Lyman limit systems, the damped Ly α systems, the effective opacity of Ly α , and the mean free path $\lambda_{\text{mfp}}^{912}$. Prochaska et al (2014) have adopted a cubic Hermite spline model (constrained to decrease monotonically) to describe $f(N_{\text{HI}})$. This model is not physical, i.e. it is purely mathematical, but it well describes the data and is smoother than broken power-laws.

With MCMC techniques they derived the curve shown in Figure 49. Perhaps the most obvious feature in $f(N_{\text{HI}})$ is the inflection at $N_{\text{HI}} \approx 10^{18} \text{ cm}^{-2}$. Such an inflection was predicted from theory (Zheng and Miralda-Escudé, 2002; Altay et al, 2011), due to the transition from optically thin to thick HI gas.

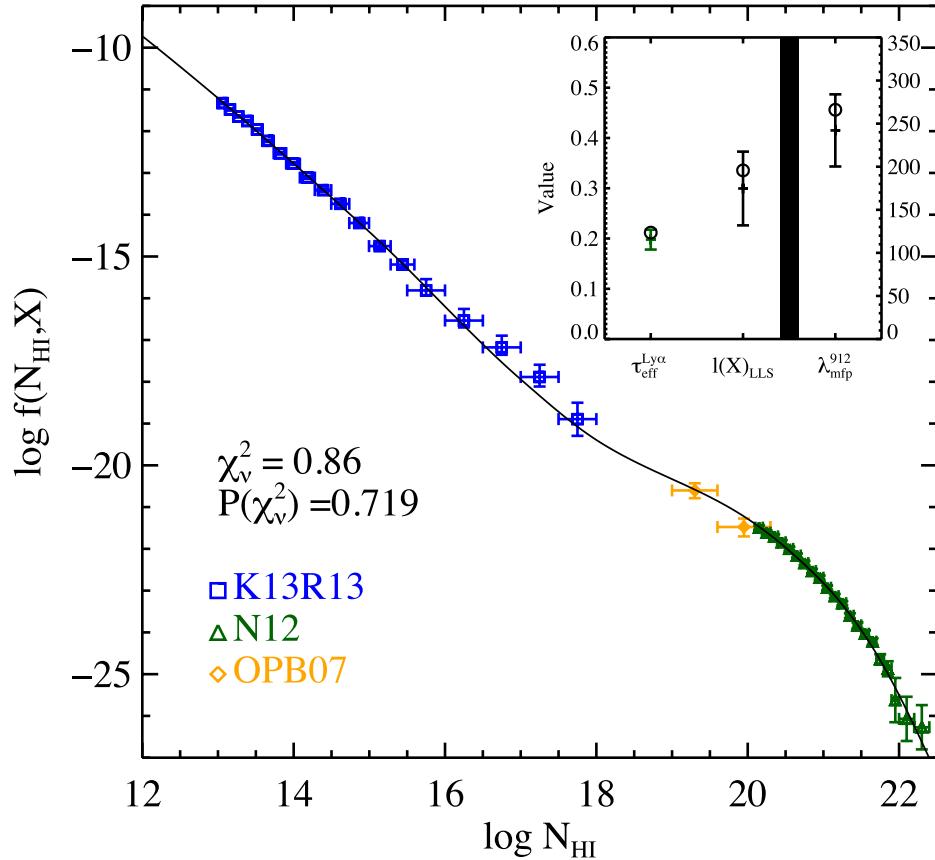


Fig. 49 Complete estimation of $f(N_{\text{HI}})$ for gas in the IGM ($N_{\text{HI}} < 10^{15} \text{ cm}^{-2}$), optically thick gas, and the damped Ly α systems. The black curve is a Hermite spline fit to the data using MCMC techniques. The analysis includes integral constraints from the effective Ly α opacity, the incidence of LLS, and estimations of the mean free path. Taken from Prochaska et al (2014).

Acknowledgements The author acknowledges the efforts of the IGM community which he has attempted to summarize in this Chapter. Without a doubt, however, many important additional works were not recognized.

References

- Altay G, Theuns T, Schaye J, Crighton NHM, Dalla Vecchia C (2011) Through Thick and Thin: H I Absorption in Cosmological Simulations. *ApJL*737:L37, DOI 10.1088/2041-8205/737/2/L37, 1012.4014

Bahcall JN (1967) Phenomenological Limits on the Absorbing Regions of Quasi-Stellar Sources. *ApJL*149:L7, DOI 10.1086/180041

Bahcall JN, Peebles PJE (1969) Statistical Tests for the Origin of Absorption Lines Observed in Quasi-Stellar Sources. *ApJL*156:L7+

Bahcall JN, Salpeter EE (1966) Absorption Lines in the Spectra of Distant Sources. *ApJ*144:847, DOI 10.1086/148675

Bahcall JN, Spitzer LJ (1969) Absorption Lines Produced by Galactic Halos. *ApJL*156:L63, URL <http://>

Bahcall JN, Wolf RA (1968) Fine-Structure Transitions. *ApJ*152:701–+

Bajtlik S, Duncan RC, Ostriker JP (1988) Quasar ionization of Lyman-alpha clouds - The proximity effect, a probe of the ultraviolet background at high redshift. *ApJ*327:570–583, DOI 10.1086/166217

Becker GD, Hewett PC, Worseck G, Prochaska JX (2013) A refined measurement of the mean transmitted flux in the Ly α forest over $2 < z < 5$ using composite quasar spectra. *MNRAS*430:2067–2081, DOI 10.1093/mnras/stt031, 1208.2584

Bergeron J, Boisse P (1991) Properties of the galaxies giving rise to MgII quasar absorption systems. *Advances in Space Research* 11:241–244, DOI 10.1016/0273-1177(91)90496-7

Boksenberg A, Carswell RF, Smith MG, Whelan JA (1978) The absorption-line spectrum of Q 1246 - 057. *MNRAS*184:773–782, DOI 10.1093/mnras/184.4.773

Braun R (2012) Cosmological Evolution of Atomic Gas and Implications for 21 cm H I Absorption. *ApJ*749:87, DOI 10.1088/0004-637X/749/1/87, 1202.1840

Brown RL, Roberts MS (1973) 21-CENTIMETER Absorption at $z = 0.692$ in the Quasar 3c 286. *ApJL*184:L7, DOI 10.1086/181276

Burbidge EM, Lynds CR, Burbidge GR (1966) On the Measurement and Interpretation of Absorption Features in the Spectrum of the Quasi-Stellar Object 3c 191. *ApJ*144:447, DOI 10.1086/148629

Busca NG, Delubac T, Rich J, Bailey S, Font-Ribera A, Kirkby D, Le Goff JM, Pieri MM, Slosar A, Aubourg É, Bautista JE, Bizyaev D, Blomqvist M, Bolton AS, Bovy J, Brewington H, Borde A, Brinkmann J, Carithers B, Croft RAC, Dawson KS, Ebelke G, Eisenstein DJ, Hamilton JC, Ho S, Hogg DW, Honscheid K, Lee KG, Lundgren B, Malanushenko E, Malanushenko V, Margala D, Maraston C, Mehta K, Miralda-Escudé J, Myers AD, Nichol RC, Noterdaeme P, Olmstead MD, Oravetz D, Palanque-Delabrouille N, Pan K, Pâris I, Percival WJ, Pe-
pe

- titjean P, Roe NA, Rollinde E, Ross NP, Rossi G, Schlegel DJ, Schneider DP, Shelden A, Sheldon ES, Simmons A, Snedden S, Tinker JL, Viel M, Weaver BA, Weinberg DH, White M, Yèche C, York DG (2013) Baryon acoustic oscillations in the Ly α forest of BOSS quasars. *A&A*552:A96, DOI 10.1051/0004-6361/201220724, 1211.2616
- Cooksey KL, Thom C, Prochaska JX, Chen H (2010) The Last Eight-Billion Years of Intergalactic C IV Evolution. *ApJ*708:868–908, DOI 10.1088/0004-637X/708/1/868, 0906.3347
- Davé R, Tripp TM (2001) The Statistical and Physical Properties of the Low-Redshift LY α Forest Observed with the Hubble Space Telescope/STIS. *ApJ*553:528–537, DOI 10.1086/320977, arXiv:astro-ph/0101419
- Dekel A, Birnboim Y (2006) Galaxy bimodality due to cold flows and shock heating. *MNRAS*368:2–20, DOI 10.1111/j.1365-2966.2006.10145.x, arXiv:astro-ph/0412300
- Duncan RC, Ostriker JP, Bajtlik S (1989) Voids in the Ly-alpha forest. *ApJ*345:39–51, DOI 10.1086/167879
- Fall SM, Pei YC (1993) Obscuration of quasars by dust in damped Lyman-alpha systems. *ApJ*402:479–492, DOI 10.1086/172151
- Faucher-Giguère CA, Kereš D (2011) The small covering factor of cold accretion streams. *MNRAS*412:L118–L122, DOI 10.1111/j.1745-3933.2011.01018.x, 1011.1693
- Faucher-Giguère CA, Lidz A, Hernquist L, Zaldarriaga M (2008a) A Flat Photoionization Rate at $2 \leq z \leq 4.2$: Evidence for a Stellar-Dominated UV Background and against a Decline of Cosmic Star Formation beyond $z \sim 3$. *ApJL*682:L9, DOI 10.1086/590409, 0806.0372
- Faucher-Giguère CA, Prochaska JX, Lidz A, Hernquist L, Zaldarriaga M (2008b) A Direct Precision Measurement of the Intergalactic Ly α Opacity at $2 \leq z \leq 4.2$. *ApJ*681:831–855, DOI 10.1086/588648, arXiv:0709.2382
- Faucher-Giguère CA, Lidz A, Zaldarriaga M, Hernquist L (2009) A New Calculation of the Ionizing Background Spectrum and the Effects of He II Reionization. *ApJ*703:1416–1443, DOI 10.1088/0004-637X/703/2/1416, 0901.4554
- Field GB (1959) An Attempt to Observe Neutral Hydrogen Between the Galaxies. *ApJ*129:525, DOI 10.1086/146652
- Font-Ribera A, Miralda-Escudé J, Arnau E, Carithers B, Lee KG, Noterdaeme P, Pâris I, Petitjean P, Rich J, Rollinde E, Ross NP, Schneider DP, White M, York DG (2012) The large-scale cross-correlation of Damped Lyman alpha systems with the Lyman alpha forest: first measurements from BOSS. *JCAP*11:059, DOI 10.1088/1475-7516/2012/11/059, 1209.4596
- Fumagalli M, Prochaska JX, Kasen D, Dekel A, Ceverino D, Primack JR (2011) Absorption-line systems in simulated galaxies fed by cold streams. *MNRAS*418:1796–1821, DOI 10.1111/j.1365-2966.2011.19599.x, 1103.2130
- Fumagalli M, O'Meara JM, Prochaska JX, Worseck G (2013) Dissecting the Properties of Optically Thick Hydrogen at the Peak of Cosmic Star Formation History. *ApJ*775:78, DOI 10.1088/0004-637X/775/1/78, 1308.1101

- Fynbo JPU, Jakobsson P, Prochaska JX, Malesani D, Ledoux C, de Ugarte Postigo A, Nardini M, Vreeswijk PM, Wiersema K, Hjorth J, Sollerman J, Chen H, Thöne CC, Björnsson G, Bloom JS, Castro-Tirado AJ, Christensen L, De Cia A, Fruchter AS, Gorosabel J, Graham JF, Jaunsen AO, Jensen BL, Kann DA, Kouveliotou C, Levan AJ, Maund J, Masetti N, Milvang-Jensen B, Palazzi E, Perley DA, Pian E, Rol E, Schady P, Starling RLC, Tanvir NR, Watson DJ, Xu D, Augusteijn T, Grundahl F, Telting J, Quirion P (2009) Low-resolution Spectroscopy of Gamma-ray Burst Optical Afterglows: Biases in the Swift Sample and Characterization of the Absorbers. *ApJS185*:526–573, DOI 10.1088/0067-0049/185/2/526, 0907.3449
- Gardner JP, Katz N, Hernquist L, Weinberg DH (2001) Simulations of Damped Ly α and Lyman Limit Absorbers in Different Cosmologies: Implications for Structure Formation at High Redshift. *ApJ559*:131–146, DOI 10.1086/322403
- Giallongo E, Grazian A, Fiore F, Fontana A, Pentericci L, Vanzella E, Dickinson M, Kocevski D, Castellano M, Cristiani S, Ferguson H, Finkelstein S, Grogan N, Hathi N, Koekemoer AM, Newman JA, Salvato M (2015) Faint AGNs at $z > 4$ in the CANDELS GOODS-S field: looking for contributors to the reionization of the Universe. *A&A578*:A83, DOI 10.1051/0004-6361/201425334, 1502.02562
- Gunn JE, Peterson BA (1965) On the Density of Neutral Hydrogen in Intergalactic Space. *ApJ142*:1633–1641, DOI 10.1086/148444
- Haardt F, Madau P (1996) Radiative Transfer in a Clumpy Universe. II. The Ultraviolet Extragalactic Background. *ApJ461*:20–+, DOI 10.1086/177035, astro-ph/9509093
- Hui L, Gnedin NY (1997) Equation of state of the photoionized intergalactic medium. *MNRAS292*:27–+, arXiv:astro-ph/9612232
- Hui L, Rutledge RE (1999) The B Distribution and the Velocity Structure of Absorption Peaks in the LYalpha Forest. *ApJ517*:541–548, DOI 10.1086/307202, arXiv:astro-ph/9709100
- Inoue AK, Shimizu I, Iwata I, Tanaka M (2014) An updated analytic model for attenuation by the intergalactic medium. *MNRAS442*:1805–1820, DOI 10.1093/mnras/stu936, 1402.0677
- Kereš D, Katz N, Fardal M, Davé R, Weinberg DH (2009) Galaxies in a simulated Λ CDM Universe - I. Cold mode and hot cores. *MNRAS395*:160–179, DOI 10.1111/j.1365-2966.2009.14541.x, 0809.1430
- Kim TS, Partl AM, Carswell RF, Müller V (2013) The evolution of H I and C IV quasar absorption line systems at $1.9 < z < 3.2$. *A&A552*:A77, DOI 10.1051/0004-6361/201220042, 1302.6622
- Kirkman D, Tytler D (1997) Intrinsic Properties of the $z = 2.7$ LY alpha Forest from Keck Spectra of Quasar HS 1946+7658. *ApJ484*:672–+, DOI 10.1086/304371, astro-ph/9701209
- Kirkman D, Tytler D, Suzuki N, Melis C, Hollywood S, James K, So G, Lubin D, Jena T, Norman ML, Paschos P (2005) The HI opacity of the intergalactic medium at redshifts $1.6 < z < 3.2$. *MNRAS360*:1373–1380, DOI 10.1111/j.1365-2966.2005.09126.x, arXiv:astro-ph/0504391
- Lee HW (2003) Asymmetric Deviation of the Scattering Cross Section around Ly α by Atomic Hydrogen. *ApJ594*:637–641, DOI 10.1086/376867, astro-ph/0308083

- Lee KG, Suzuki N, Spergel DN (2012) Mean-flux-regulated Principal Component Analysis Continuum Fitting of Sloan Digital Sky Survey Ly α Forest Spectra. AJ143:51, DOI 10.1088/0004-6256/143/2/51, 1108.6080
- Lee KG, Hennawi JF, Stark C, Prochaska JX, White M, Schlegel DJ, Eilers AC, Arinyo-i-Prats A, Suzuki N, Croft RAC, Caputi KI, Cassata P, Ilbert O, Garilli B, Koekemoer AM, Le Brun V, Le Fèvre O, Maccagni D, Nugent P, Taniguchi Y, Tasca LAM, Tresse L, Zamorani G, Zucca E (2014) Ly α Forest Tomography from Background Galaxies: The First Megaparsec-resolution Large-scale Structure Map at $z > 2$. ApJL795:L12, DOI 10.1088/2041-8205/795/1/L12, 1409.5632
- Lee KG, Hennawi JF, Spergel DN, Weinberg DH, Hogg DW, Viel M, Bolton JS, Bailey S, Pieri MM, Carithers W, Schlegel DJ, Lundgren B, Palanque-Delabrouille N, Suzuki N, Schneider DP, Yèche C (2015) IGM Constraints from the SDSS-III/BOSS DR9 Ly α Forest Transmission Probability Distribution Function. ApJ799:196, DOI 10.1088/0004-637X/799/2/196, 1405.1072
- López S, D’Odorico V, Ellison SL, Becker GD, Christensen L, Cupani G, Denney KD, Pâris I, Worseck G, Berg TAM, Cristiani S, Dessauges-Zavadsky M, Haehnelt M, Hamann F, Hennawi J, Iršič V, Kim TS, López P, Lund Saust R, Ménard B, Perrotta S, Prochaska JX, Sánchez-Ramírez R, Vestergaard M, Viel M, Wisotzki L (2016) XQ-100: A legacy survey of one hundred $3.5 < z < 4.5$ quasars observed with VLT/X-shooter. A&A594:A91, DOI 10.1051/0004-6361/201628161, 1607.08776
- Lowenthal JD, Koo DC, Guzman R, Gallego J, Phillips AC, Faber SM, Vogt NP, Illingworth GD, Gronwall C (1997) Keck Spectroscopy of Redshift Z approximately 3 Galaxies in the Hubble Deep Field. ApJ481:673–+, DOI 10.1086/304092, arXiv:astro-ph/9612239
- Lusso E, Comastri A, Simmons BD, Mignoli M, Zamorani G, Vignali C, Brusa M, Shankar F, Lutz D, Trump JR, Maiolino R, Gilli R, Bolzonella M, Puccetti S, Salvato M, Impey CD, Civano F, Elvis M, Mainieri V, Silverman JD, Koekemoer AM, Bongiorno A, Merloni A, Berta S, Le Floc’h E, Magnelli B, Pozzi F, Riguccini L (2012) Bolometric luminosities and Eddington ratios of X-ray selected active galactic nuclei in the XMM-COSMOS survey. MNRAS425:623–640, DOI 10.1111/j.1365-2966.2012.21513.x, 1206.2642
- Lusso E, Worseck G, Hennawi JF, Prochaska JX, Vignali C, Stern J, O’Meara JM (2015) The first ultraviolet quasar-stacked spectrum at $z \sim 2.4$ from WFC3. MNRAS449:4204–4220, DOI 10.1093/mnras/stv516, 1503.02075
- Madau P (1995) Radiative transfer in a clumpy universe: The colors of high-redshift galaxies. ApJ441:18–27, URL <http://>
- Madau P, Haardt F (2015) Cosmic Reionization after Planck: Could Quasars Do It All? ApJL813:L8, DOI 10.1088/2041-8205/813/1/L8, 1507.07678
- McQuinn M (2016) The Evolution of the Intergalactic Medium. AAR&A54:313–362, DOI 10.1146/annurev-astro-082214-122355, 1512.00086
- Meiksin A, Madau P (1993) On the photoionization of the intergalactic medium by quasars at high redshift. ApJ412:34–55, DOI 10.1086/172898
- Meiksin AA (2009) The physics of the intergalactic medium. Reviews of Modern Physics 81:1405–1469, DOI 10.1103/RevModPhys.81.1405, 0711.3358

- Miralda-Escudé J, Cen R, Ostriker JP, Rauch M (1996) The Ly alpha Forest from Gravitational Collapse in the Cold Dark Matter + Lambda Model. *ApJ*471:582–+, DOI 10.1086/177992, arXiv:astro-ph/9511013
- Miralda-Escudé J, Haehnelt M, Rees MJ (2000) Reionization of the Inhomogeneous Universe. *ApJ*530:1–16, DOI 10.1086/308330, astro-ph/9812306
- Møller P, Jakobsen P (1990) The Lyman continuum opacity at high redshifts - Through the Lyman forest and beyond the Lyman valley. *A&A*228:299–309
- Nestor DB, Shapley AE, Kornei KA, Steidel CC, Siana B (2013) A Refined Estimate of the Ionizing Emissivity from Galaxies at $z \approx 3$: Spectroscopic Follow-up in the SSA22a Field. *ApJ*765:47, DOI 10.1088/0004-637X/765/1/47, 1210.2393
- Noterdaeme P, Petitjean P, Carithers WC, Pâris I, Font-Ribera A, Bailey S, Aubourg E, Bizyaev D, Ebelke G, Finley H, Ge J, Malanushenko E, Malanushenko V, Miralda-Escudé J, Myers AD, Oravetz D, Pan K, Pieri MM, Ross NP, Schneider DP, Simmons A, York DG (2012) Column density distribution and cosmological mass density of neutral gas: Sloan Digital Sky Survey-III Data Release 9. *A&A*547:L1, DOI 10.1051/0004-6361/201220259, 1210.1213
- O'Meara JM, Prochaska JX, Worseck G, Chen HW, Madau P (2013) The HST/ACS+WFC3 Survey for Lyman Limit Systems. II. Science. *ApJ*765:137, DOI 10.1088/0004-637X/765/2/137, 1204.3093
- O'Meara JM, Lehner N, Howk JC, Prochaska JX, Fox AJ, Swain MA, Gelino CR, Berriman GB, Tran H (2015) The First Data Release of the KODIAQ Survey. *AJ*150:111, DOI 10.1088/0004-6256/150/4/111, 1505.03529
- Ostriker JP, Cowie LL (1981) Galaxy formation in an intergalactic medium dominated by explosions. *ApJL*243:L127–L131, DOI 10.1086/183458
- Ostriker JP, Heisler J (1984) Are cosmologically distant objects obscured by dust? - A test using quasars. *ApJ*278:1–10, DOI 10.1086/161762
- Ostriker JP, Ikeuchi S (1983) Physical properties of the intergalactic medium and the Lyman-alpha absorbing clouds. *ApJL*268:L63–L68, DOI 10.1086/184030
- Pâris I, Petitjean P, Rollinde E, Aubourg E, Busca N, Charlassier R, Delubac T, Hamilton JC, Le Goff JM, Palanque-Delabrouille N, Peirani S, Pichon C, Rich J, Vargas-Magaña M, Yèche C (2011) A principal component analysis of quasar UV spectra at $z \sim 3$. *A&A*530:A50, DOI 10.1051/0004-6361/201016233, 1104.2024
- Penton SV, Shull JM, Stocke JT (2000) The Local Ly α Forest. II. Distribution of H I Absorbers, Doppler Widths, and Baryon Content. *ApJ*544:150–175, DOI 10.1086/317179, astro-ph/9911128
- Petitjean P, Webb JK, Rauch M, Carswell RF, Lanzetta K (1993) Evidence for structure in the H I column density distribution of QSO absorbers. *MNRAS*262:499–505
- Press WH, Rybicki GB, Schneider DP (1993) Properties of high-redshift Lyman-alpha clouds. I - Statistical analysis of the Schneider-Schmidt-Gunn quasars. *ApJ*414:64–81, DOI 10.1086/173057, astro-ph/9303016
- Prochaska JX, Wolfe AM (2009) On the (Non)Evolution of H I Gas in Galaxies Over Cosmic Time. *ApJ*696:1543–1547, DOI 10.1088/0004-637X/696/2/1543, 0811.2003

- Prochaska JX, Herbert-Fort S, Wolfe AM (2005) The SDSS Damped Ly α Survey: Data Release 3. *ApJ*635:123–142, DOI 10.1086/497287
- Prochaska JX, Worseck G, O’Meara JM (2009) A Direct Measurement of the Intergalactic Medium Opacity to H I Ionizing Photons. *ApJL*705:L113–L117, DOI 10.1088/0004-637X/705/2/L113, 0910.0009
- Prochaska JX, O’Meara JM, Worseck G (2010) A Definitive Survey for Lyman Limit Systems at $z \sim 3.5$ with the Sloan Digital Sky Survey. *ApJ*718:392–416, DOI 10.1088/0004-637X/718/1/392, 0912.0292
- Prochaska JX, Hennawi JF, Lee KG, Cantalupo S, Bovy J, Djorgovski SG, Ellison SL, Wingyee Lau M, Martin CL, Myers A, Rubin KHR, Simcoe RA (2013a) Quasars Probing Quasars. VI. Excess H I Absorption within One Proper Mpc of $z \sim 2$ Quasars. *ApJ*776:136, DOI 10.1088/0004-637X/776/2/136, 1308.6222
- Prochaska JX, Hennawi JF, Simcoe RA (2013b) A Substantial Mass of Cool, Metal-enriched Gas Surrounding the Progenitors of Modern-day Ellipticals. *ApJL*762:L19, DOI 10.1088/2041-8205/762/2/L19, 1211.6131
- Prochaska JX, Madau P, O’Meara JM, Fumagalli M (2014) Towards a unified description of the intergalactic medium at redshift $z \sim 2.5$. *MNRAS*438:476–486, DOI 10.1093/mnras/stt2218, 1310.0052
- Prochaska JX, Werk JK, Worseck G, Tripp TM, Tumlinson J, Burchett JN, Fox AJ, Fumagalli M, Lehner N, Peebles MS, Tejos N (2017) The COS-Halos Survey: Metallicities in the Low-redshift Circumgalactic Medium. *ApJ*837:169, DOI 10.3847/1538-4357/aa6007, 1702.02618
- Rahmati A, Pawlik AH, Raicevic M, Schaye J (2013) On the evolution of the H I column density distribution in cosmological simulations. *MNRAS*430:2427–2445, DOI 10.1093/mnras/stt066, 1210.7808
- Ribaudo J, Lehner N, Howk JC (2011) A Hubble Space Telescope Study of Lyman Limit Systems: Census and Evolution. *ApJ*736:42–+, DOI 10.1088/0004-637X/736/1/42, 1105.0659
- Richards GT, Lacy M, Storrie-Lombardi LJ, Hall PB, Gallagher SC, Hines DC, Fan X, Papovich C, Vanden Berk DE, Trammell GB, Schneider DP, Vestergaard M, York DG, Jester S, Anderson SF, Budavári T, Szalay AS (2006) Spectral Energy Distributions and Multiwavelength Selection of Type 1 Quasars. *ApJS*166:470–497, DOI 10.1086/506525, arXiv:astro-ph/0601558
- Robertson JG (2013) Quantifying Resolving Power in Astronomical Spectra. *PASA*30:e048, DOI 10.1017/pasa.2013.26, 1308.0871
- Rudie GC, Steidel CC, Shapley AE, Pettini M (2013) The Column Density Distribution and Continuum Opacity of the Intergalactic and Circumgalactic Medium at Redshift $z = 2.4$. *ApJ*769:146, DOI 10.1088/0004-637X/769/2/146, 1304.6719
- Sargent WLW, Young PJ, Boksenberg A, Tytler D (1980) The distribution of Lyman-alpha absorption lines in the spectra of six QSOs - Evidence for an intergalactic origin. *ApJS*42:41–81, DOI 10.1086/190644
- Sargent WLW, Steidel CC, Boksenberg A (1989) A survey of Lyman-limit absorption in the spectra of 59 high-redshift QSOs. *ApJS*69:703–761, DOI 10.1086/191326

- Schmidt M (1963) 3C 273: a star-like object with large red-shift. *Nature*197:1040–1040
- Songaila A, Cowie LL (2002) Approaching Reionization: The Evolution of the Ly α Forest from z=4 to z=6. *AJ*123:2183–2196, DOI 10.1086/340079, astro-ph/0202165
- Steidel CC, Giavalisco M, Pettini M, Dickinson M, Adelberger KL (1996) Spectroscopic Confirmation of a Population of Normal Star-forming Galaxies at Redshifts $z > 3$. *ApJL*462:L17+, DOI 10.1086/310029, arXiv:astro-ph/9602024
- Storrie-Lombardi LJ, Irwin MJ, McMahon RG (1996) APM $z > 4$ survey: distribution and evolution of high column density HI absorbers. *MNRAS*282:1330–1342, arXiv:astro-ph/9608146
- Suzuki N (2006) Quasar Spectrum Classification with Principal Component Analysis (PCA): Emission Lines in the Ly α Forest. *ApJS*163:110–121, DOI 10.1086/499272
- Telfer RC, Zheng W, Kriss GA, Davidsen AF (2002) The Rest-Frame Extreme-Ultraviolet Spectral Properties of Quasi-stellar Objects. *ApJ*565:773–785, DOI 10.1086/324689, astro-ph/0109531
- Tytler D (1982) QSO Lyman limit absorption. *Nature*298:427–432, DOI 10.1038/298427a0
- Tytler D (1987) The distribution of QSO absorption system column densities - Evidence for a single population. *ApJ*321:49–68, DOI 10.1086/165615
- van de Voort F, Schaye J (2012) Properties of gas in and around galaxy haloes. *MNRAS*423:2991–3010, DOI 10.1111/j.1365-2966.2012.20949.x, 1111.5039
- Vanden Berk DE, Richards GT, Bauer A, Strauss MA, Schneider DP, Heckman TM, York DG, Hall PB, Fan X, Knapp GR, Anderson SF, Annis J, Bahcall NA, Bernardi M, Briggs JW, Brinkmann J, Brunner R, Burles S, Carey L, Castander FJ, Connolly AJ, Crocker JH, Csabai I, Doi M, Finkbeiner D, Friedman S, Friedman JA, Fukugita M, Gunn JE, Hennessy GS, Ivezić Ž, Kent S, Kunszt PZ, Lamb DQ, Leger RF, Long DC, Loveday J, Lupton RH, Meiksin A, Merelli A, Munn JA, Newberg HJ, Newcomb M, Nichol RC, Owen R, Pier JR, Pope A, Rockosi CM, Schlegel DJ, Siegmund WA, Smee S, Snir Y, Stoughton C, Stubbs C, SubbaRao M, Szalay AS, Szokoly GP, Tremonti C, Uomoto A, Waddell P, Yanny B, Zheng W (2001) Composite Quasar Spectra from the Sloan Digital Sky Survey. *AJ*122:549–564, DOI 10.1086/321167
- Vogt SS, Allen SL, Bigelow BC, Bresee L, Brown B, Cantrall T, Conrad A, Couture M, Delaney C, Epps HW, Hilyard D, Hilyard DF, Horn E, Jern N, Kanto D, Keane MJ, Kibrick RI, Lewis JW, Osborne J, Pardeilhan GH, Pfister T, Ricketts T, Robinson LB, Stover RJ, Tucker D, Ward J, Wei MZ (1994) HIRES: the high-resolution echelle spectrometer on the Keck 10-m Telescope. In: Proc. SPIE Instrumentation in Astronomy VIII, David L. Crawford; Eric R. Craine; Eds., Volume 2198, p. 362, pp 362–+
- Wolfe AM, Turnshek DA, Smith HE, Cohen RD (1986) Damped Lyman-alpha absorption by disk galaxies with large redshifts. I - The Lick survey. *ApJS*61:249–304, DOI 10.1086/191114

- Wolfe AM, Lanzetta KM, Foltz CB, Chaffee FH (1995) The Large Bright QSO Survey for Damped LY alpha Absorption Systems. *ApJ*454:698–+, DOI 10.1086/176523
- Wolfe AM, Gawiser E, Prochaska JX (2005) DAMPED LYA SYSTEMS. *A&A*43:861–918
- Worseck G, Prochaska JX, O'Meara JM, Becker GD, Ellison SL, Lopez S, Meiksin A, Ménard B, Murphy MT, Fumagalli M (2014) The Giant Gemini GMOS survey of $z_{em} > 4.4$ quasars - I. Measuring the mean free path across cosmic time. *MNRAS*445:1745–1760, DOI 10.1093/mnras/stu1827, 1402.4154
- Young PJ, Sargent WLW, Boksenberg A, Carswell RF, Whelan JA (1979) A high-resolution study of the absorption spectrum of PKS 2126-158. *ApJ*229:891–908, DOI 10.1086/157024
- Zheng Z, Miralda-Escudé J (2002) Self-shielding Effects on the Column Density Distribution of Damped Ly α Systems. *ApJL*568:L71–L74, DOI 10.1086/340330, astro-ph/0201275
- Zwaan MA, van der Hulst JM, Briggs FH, Verheijen MAW, Ryan-Weber EV (2005) Reconciling the local galaxy population with damped Lyman α cross-sections and metal abundances. *MNRAS*364:1467–1487, DOI 10.1111/j.1365-2966.2005.09698.x, arXiv:astro-ph/0510127