Astronomy & Astrophysics manuscript no. output
February 2, 2018

Neutral gas properties of Lyman continuum emitting galaxies: column densities and covering fractions from UV absorption lines

S. Gazagnes[1,2,3,4], J. Chisholm[1], D. Schaerer[1,5], A. Verhamme[1], J. R. Rigby[6], and M. Bayliss[7]

[1] Observatoire de Genève, Université de Genève, 51 Ch. des Maillettes, 1290 Versoix, Switzerland
[2] Johan Bernouilli Institute, University of Groningen, P.O Box 407, 9700 Groningen, AK, The Netherlands
[3] Kapteyn Astronomical Institute, University of Groningen, P.O Box 800, 9700 AV Groningen, The Netherlands
[4] KVI-Center for Advanced Radiation Technology (KVI-CART), Zernikelaan 25, Groningen 9747 AA, The Netherlands
[5] CNRS, IRAP, 14 Avenue E. Belin, 31400 Toulouse, France
[6] Observational Cosmology Lab, NASA Goddard Space Flight Center, 8800 Greenbelt Rd., Greenbelt, MD 20771, USA
[7] MIT Kavli Institute for Astrophysics and Space Research, 77 Massachusetts Ave., Cambridge, MA 02139, USA

Received <date> / Accepted <date>

ABSTRACT

Context. The processes allowing the escape of ionizing photons from galaxies into the intergalactic medium are poorly known.
Aims. To understand how Lyman continuum (LyC) photons escape galaxies, we constrain the H I covering fractions and column densities using ultraviolet H I and metal absorption lines of 18 star-forming galaxies which have Lyman series observations. Nine of these galaxies are confirmed LyC emitters.
Methods. We fit the stellar continuum, dust attenuation, metal, and H I properties to consistently determine the UV attenuation, as well as the column densities and covering factors of neutral hydrogen and metals. We use synthetic interstellar absorption lines to explore the systematics of our measurements. Then we apply our method to the observed UV spectra of low-redshift and $z \sim 2$ galaxies.
Results. The observed H I lines are found to be saturated in all galaxies. An indirect approach using O I column densities and the observed O/H abundances yields H I column densities of $\log(N_{\rm H\,I}) \sim 18.6 - 20$ cm^{-2}. These columns are too high to allow the escape of ionizing photons. We find that the known LyC leakers have H I covering fractions less than unity. Ionizing photons escape through optically thin holes/channels in a clumpy interstellar medium. Our simulations confirm that the H I covering fractions are accurately recovered. The Si II and H I covering fractions scale linearly, in agreement with observations from stacked Lyman break galaxy spectra at $z \sim 3$. Thus, with an empirical correction, the Si II absorption lines can also be used to determine the H I coverage. Finally, we show that a consistent fitting of dust attenuation, continuum and absorption lines is required to properly infer the covering fraction of neutral gas and subsequently to infer the escape fraction of ionizing radiation.
Conclusions. These measurements can estimate the LyC escape fraction, as we demonstrate in a companion paper.

Key words. galaxies: ISM – ISM: abundances – ISM: lines and bands – Ultraviolet: ISM – dust, extinction – dark ages, reionization, first stars

1. Introduction

Star-forming galaxies are ideal laboratories to understand how the early universe became reionized. Galaxies likely reionized the universe because quasars are too rare at high redshifts (Fontanot et al. 2012, 2014). Compact galaxies with intense star formation rates produce large amounts of ionizing photons which, under certain circumstances, escape the interstellar medium (ISM) and ionize the intergalactic medium (IGM). To reionize the universe, studies suggest that 10 – 20% of the ionizing photons produced by star-forming galaxies must escape galaxies (Ouchi et al. 2009; Robertson et al. 2013; Dressler et al. 2015). However, it has been challenging to detect Lyman continuum radiation from individual galaxies.

Three unambiguous observations of ionizing photons have been reported at z~3 (Vanzella et al. 2015; de Barros et al. 2016; Shapley et al. 2016; Bian et al. 2017), with Lyman continuum escape fractions (f_{esc}^{LyC}) greater than the escape fractions required to reionize the universe (near 50%). Additionally, there are nine low-redshift (z < 0.3) galaxies with $1\% \leq f_{esc}^{LyC} \leq 13\%$ (Bergvall et al. 2006; Leitet et al. 2013; Borthakur et al. 2014; Leitherer et al. 2016; Izotov et al. 2016b,a; Puschnig et al. 2017), and one recent detection at z = 0.37 with $f_{esc}^{LyC} \approx 46\%$ (Izotov et al. 2018). The low number of detections emphasizes the difficulty of detecting Lyman continuum Emitters (LCEs).

Zackrisson et al. (2013) proposed two theoretical models to explain how ionizing photons escape galaxies. In the first scenario, low H I column densities allow Lyman continuum photons to pass through without being completely absorbed; this is called the density-bounded scenario (Jaskot & Oey 2013; Nakajima & Ouchi 2014). This scenario manifests itself as low H I column densities ($< 10^{18}$ cm^{-2}). In the second scenario, ionizing photons leak into the IGM through holes in the neutral gas (Heckman et al. 2011). This scenario is called the picket-fence model and relies on a patchy neutral gas. A patchy neutral ISM manifests as H I absorption lines with a covering fraction less than one. It is unclear which of these scenarios describes how ionizing photons leak from galaxies. Constraining neutral gas properties, especially the H I covering fraction and column density, is an effective way to disentangle how ionizing photons escape galaxies.

Send offprint requests to: s.r.n.gazagnes@rug.nl

H I absorption lines in the restframe far ultraviolet (the Lyman series: 912-1026Å) directly probe the H I covering fraction and column density. However, the Lyman series is challenging to observe for several reasons. First, at low-redshifts it requires deep restframe far-ultraviolet observations blueward of Lyα, a wavelength regime notoriously difficult to observe. Second, the Lyman series is unavailable at redshift z > 3 because the Lyα forest completely absorbs this region. Third, several observational obstacles need to be accounted for to measure the H I properties. In particular, foreground contamination (Vanzella et al. 2010), intervening absorbers, stellar continuum from the galaxy itself, ISM absorption lines, and, at low-redshifts, Milky Way and geocoronal emission need to be identified.

ISM metal absorption lines (i.e. Si II 1260Å, C II 1334Å) are easier to observe than the Lyman series (Heckman et al. 2011; Alexandroff et al. 2015). However, using metal absorption lines to trace the H I assumes that metals directly probe the neutral gas. Recent studies indicate that the ISM metal lines may have a factor of 2 times smaller covering fractions than H I absorption lines (Reddy et al. 2016b). As a result, metal absorption lines may not trace the H I.

In this article, we directly observe the Lyman series of individual high and low-redshift star-forming galaxies to determine their neutral gas properties, and to compare H I measurements to ISM metal properties. For the first time, we measure the H I properties of spectroscopically confirmed LyC emitters to determine which physical process enables ionizing photons to escape galaxies. A companion paper uses these observed H I properties to predict the escape fractions of ionizing photons (Chisholm et al. submitted; hereafter Paper II).

This paper is organized as follows: Sect. 2 describes the observational data. Sect. 3 defines the different models and equations used to fit the stellar continua and UV absorption lines. In Sect. 4 we use synthetic spectra to illustrate how accurately we recover H I column densities and covering fractions from observations. Sect. 5 discusses the measured H I covering fractions of the LyC emitters, the relation between the H I and Si II covering fractions, the effects of the assumed dust geometry, and comparisons to previous studies. We summarize our results in Sect. 6.

2. Observed data

We study the neutral gas properties of a sample of 18 star-forming galaxies listed in Table 1. Our selection is driven by the need to observe the Lyman series, i.e. available restframe UV spectroscopy between Lyman-β and the Lyman limit. We select the low-redshift galaxies observed with the Cosmic Origins Spectrograph (COS) on the *Hubble Space Telescope* (Green et al. 2012) with this wavelength coverage. Given the sensitivity and wavelength range of the G130M grating on COS, the Lyman series is observable with a spectral resolution $R \gtrsim 1500$ for galaxies at $z > 0.18$. Therefore, from the COS archive we selected the 15 galaxies at low redshifts ($z < 0.36$) with Lyman series observations. Each galaxy is at a redshift such that at least the Lyβ line is observable with the COS G140L or G130M gratings.

Many of our low-redshift galaxies were originally targeted to observe possible LyC emission, but only nine of them are confirmed LyC emitting galaxies. Four galaxies are from the Green Pea sample of Henry et al. (2015), and two are Lyman Break Analogs from Heckman et al. (2011). The other nine galaxies are known LyC emitters (J1503+3644, J0925+1409, J1152+3400, J1333+6246, J1442-0209, J0921+4509, Tol1247-232, Tol0440-381 and Mrk 54 from Izotov et al. 2016b,a; Borthakur et al. 2014; Leitherer et al. 2016). The nine leakers with COS/HST data were reduced using CALCOS v2.21 and a custom method for faint COS spectra (Worseck et al. 2016). The other COS/HST data were reduced with CALCOS v2.20.1 and the methods from Wakker et al. (2015). The five Izotov et al. (2016a,b) spectra were smoothed with a 5 pixel boxcar.

Table 1: Sample of galaxies with Lyman series observations

Galaxy name	z	12 + log(O/H)	f_{esc}^{LyC}	R
(1)	(2)	(3)	(4)	(5)
J0921+4509	0.23499	8.67[a]	0.010[g]	15000
J1503+3644	0.3537	7.95[b]	0.058[b]	1500
J0925+1409	0.3013	7.91[c]	0.072[c]	1500
J1152+3400	0.3419	8.00[b]	0.132[b]	1500
J1333+6246	0.3181	7.76[b]	0.056[b]	1500
J1442-0209	0.2937	7.93[b]	0.074[b]	1500
Tol1247-232	0.0482	8.10[d]	0.004[h]	1500
Tol0440-381	0.0410	8.20[d]	0.019[h]	1500
Mrk54	0.0451	8.60[d]	<0.002[h]	1500
J0926+4427	0.18069	8.01[e]	-	15000
J1429+0643	0.1736	8.20[e]	-	15000
GP0303-0759	0.16488	7.86[e]	-	15000
GP1244+0216	0.23942	8.17[e]	-	15000
GP1054+5238	0.25264	8.10[e]	-	15000
GP0911+1831	0.26223	8.00[e]	-	15000
SGAS J1226	2.92525	-	-	4000
SGAS J1527	2.76228	< 8.5[i]	-	2700
Cosmic Eye	3.07483	8.60[f]	-	2500

Notes. (1) Galaxy name; (2) redshift; (3) metallicities derived from optical emission lines; (4) Lyman continuum escape fraction; and (5) spectral resolution of the observations. Dashes indicate that the quantities have not been measured.

References. (a) Pettini & Pagel (2004); (b) Izotov et al. (2016b); (c) Izotov et al. (2016a); (d) Leitherer et al. (2016); (e) Izotov et al. (2011); (f) Stark et al. (2008); (g) Borthakur et al. (2014); (h) Chisholm et al. (2017a) (i) Wuyts et al. (2012)

Finally, we also include 3 gravitationally lensed galaxies at $z \approx 3$ (Stark et al. 2008; Koester et al. 2010). These galaxies are part of the Magellan Evolution of Galaxies Spectroscopic and Ultraviolet Reference Atlas (MegaSaura; Rigby et al. 2017), and are selected because they are the only galaxies in the MegaSaura sample with a signal-to-noise ratio greater than 2 near the Lyman series. These are moderate resolution ($R \sim 3000$) spectra observed with the MagE spectrograph on the Magellan Telescopes (Marshall et al. 2008). Instead of the full galaxy names, we use a short name for two sources SGAS J1226 = SGAS J122651.3+215220 and SGAS J1527 = SGAS J152745.1+065219.

The complete sample is summarized in Table 1, where we list a few host properties and (nominal) spectral resolutions of the observations. An upper limit on the [N II]/Hα ratio constrains the metallicity for SGAS J1527 (12+log(O/H) < 8.5; Wuyts et al. 2012); for SGAS J1226 these lines are not accessible from the ground. Meanwhile, Stark et al. (2008) measured the metallicity of the Cosmic Eye using the R_{23} index.

3. UV spectral fitting methods and results

We now describe the theory and the methods we use to fit the UV spectra including: the stellar continua, the H I absorption lines, and the metal absorption lines. The method is then applied to simulated data (in particular to determine systematic errors), and to the observed spectra.

3.1. UV continuum and interstellar absorption line modeling

3.1.1. Adopted geometries and basic formulas

To describe the radiation transfer in the host galaxy, we adopt the classical "picket-fence" model with different assumptions on the geometric distribution of gas and dust. In practice we consider two cases: *(a)* the picket-fence model with a uniform foreground dust screen, and *(b)* a clumpy picket-fence model where dust is only in the neutral gas clumps. Two parameters describe the two models: the dust attenuation (here parametrized by $E_{\mathrm{B-V}}$) and the geometric covering fraction of neutral gas (C_f), defined as the fraction of the total lines of sight of the emitted UV radiation which are intercepted by neutral gas in the direction towards the distant observer.

In *(a)* both the radiation emerging from the gas clumps (with geometric coverage C_f) and radiation directly escaping $(1 - C_f)$ is attenuated by a uniform foreground dust screen. In *(b)*, only a fraction, C_f, of radiation is "processed" through the gas clumps, imprinting interstellar absorption lines and attenuating the stellar continuum. The rest escapes unaltered. Gas within these clumps is assumed to be homogeneous, and the inter-clump medium has a negligible column density of neutral gas, i.e. is assumed to be completely transparent.

These simple models have already been examined and assumed by other authors (e.g. Zackrisson et al. 2013; Borthakur et al. 2014; Vasei et al. 2016; Reddy et al. 2016b). We therefore only briefly list the main equations used in our spectral modeling. For a picket-fence model with a uniform foreground dust screen *(a)* the emergent flux F_λ is:

$$F_\lambda = F^\star_\lambda \times 10^{-0.4 k_\lambda E_{\mathrm{B-V}}} \times \left(C_f \exp(-\tau_\lambda) + (1 - C_f)\right), \tag{1}$$

where $F^\star_\lambda$ is the intrinsic stellar emission prior to alteration by the ISM, k_λ describes the attenuation law, and τ_λ is the optical depth of the interstellar absorption lines. For a picket-fence model, with a clumpy gas distribution *(b)*, the emergent flux becomes:

$$F_\lambda = F^\star_\lambda \times 10^{-0.4 k_\lambda E_{\mathrm{B-V}}} \times C_f \exp(-\tau_\lambda) + F^\star_\lambda \times (1 - C_f), \tag{2}$$

where the second term describes the unattenuated, directly escaping, radiation. This light is unattenuated because the holes are assumed to be free of gas and dust. For large covering fractions ($C_f \rightarrow 1$) or low attenuations ($E_{\mathrm{B-V}} \rightarrow 0$), Eq. (1) and Eq. (2) are identical.

We define the residual flux, R, as the ratio of the flux density at the observed wavelength of the line to the continuum flux density. R gives the fraction of light unabsorbed by the neutral gas. For saturated lines ($\tau_\lambda \gg 1$) the residual flux becomes

$$R = 1 - C_f \tag{3}$$

for a uniform dust screen *(a)*, and

$$R = \frac{(1 - C_f)}{10^{-0.4 k_\lambda E_{\mathrm{B-V}}} C_f + (1 - C_f)} \tag{4}$$

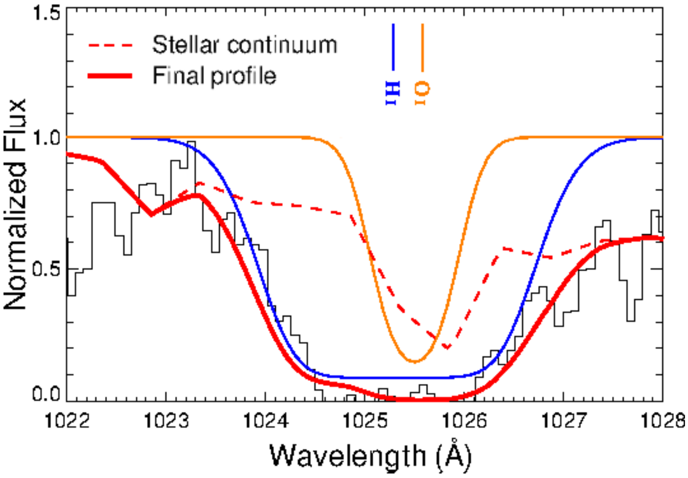

Fig. 1: Total fit (red solid line) of the Lyβ absorption line for the galaxy GP 1244+0216 (black data). The contributions from the stellar continuum (dashed red line), Lyβ absorption line (blue solid line), and O I absorption line (orange solid line) blend together. O I absorption lines at 989Å and 1039Å are unblended and robustly constrain the O I profile. All three components need to be accounted for when fitting the Lyman series.

for case *(b)* (Vasei et al. 2016). Note that R of the uniform screen model is always greater than, or equal to, the R of the clumpy model (see the discussion in Sect 5.3).

Finally, the absolute escape fraction of radiation for monochromatic radiation, $f_{\mathrm{esc}} = F_\lambda / F^\star_\lambda$, becomes

$$f_{\mathrm{esc}} = 10^{-0.4 k_\lambda E_{\mathrm{B-V}}} \times \left(C_f \exp(-\tau_\lambda) + (1 - C_f)\right) \tag{5}$$

for the uniform dust screen *(a)*, and

$$f_{\mathrm{esc}} = 10^{-0.4 k_\lambda E_{\mathrm{B-V}}} \times C_f \exp(-\tau_\lambda) + (1 - C_f) \tag{6}$$

in the clumpy geometry *(b)*. It is important to note that the values of C_f and $E_{\mathrm{B-V}}$ differ a priori between the two model geometries, except for the case of high covering fractions ($C_f \rightarrow 1$) or low attenuation ($E_{\mathrm{B-V}} \rightarrow 0$). They must be determined consistently from fits to the observed data, adopting the equations corresponding to the assumed geometry.

3.1.2. Fitting method

The main spectral region modeled here is the Lyman Series, H I absorption lines from Lyβ to the Lyman break ($\approx 912 - 1050$Å). In practice, due to the lower signal-to-noise ratio (SNR) close to the Lyman break, the bluest Lyman line that we include is Ly6 (930 Å). Figure 1 emphasizes the complicated nature of the reddest Lyman series line, Lyβ: strong stellar continuum features (red dashed line) blend with the broad H I interstellar absorption (blue line), as well as the weaker O I interstellar absorption line (orange line). Consequently, bluer Lyman series lines (especially Lyγ) have fewer complications due to their simpler stellar continua.

We model ISM metal absorption lines from O I, O VI, Si II, C II and C III (listed in Table 2). O I has a similar ionization structure as H I, such that O I absorption lines blend with all H I lines, except for O I 989 and 1039 Å. These two lines constrain the O I profile.

First, we fit the stellar continuum. We start with an initial stellar model using a linear combination of 10 single-age stellar continuum models with ages of 1, 2, 3, 4, 5, 8, 10, 15, 20

Table 2: Fitted absorption lines

Ion	$\lambda_{\rm rest}$ [Å]
H I	920.947[a]
	923.150[a]
	926.225[a]
	930.748
	937.803
	949.743
	972.536
	1025.473
O I	924.950[a]
	929.517
	930.256
	936.629
	948.685
	950.885
	971.738
	976.448
	988.578
	988.655
	988.773
	1025.762
	1039.230
	1302.168[b]
O VI	1031.926
	1037.616
C II	1036.337
C III	977.030
Si II	989.870
	1020.70
	1190.42[c]
	1193.28[c]
	1260.42[c]

Notes. Wavelengths are in vacuum.
[(a)] Used only for generating synthetic spectra, see Sect. A .
[(b)] Used for O I measurements in Tol0440-381 and Mrk54 spectra, see Sect. 3.3.2 .
[(c)] Used only to measure Si II covering fraction, see Sect.3.3.3 .

and 40 Myr. We also use stellar continuum metallicities of 0.05, 0.2, 0.4, 1 or 2 $Z_\odot$. These spectra are drawn from the fully theoretical STARBURST99 library (S99; Leitherer et al. 1999), computed with the WM-Basic method (Leitherer et al. 2010), and have a spectral resolution R(S99) ≈ 2500. We choose the stellar continuum metallicity closest to the ISM metallicity (Table 1). The STARBUST99 models use a Kroupa Initial Mass Function with a high (low) mass exponent of 2.3 (1.3), a high-mass cutoff at 100 $M_\odot$, and the stellar evolution tracks with high mass-loss from Meynet et al. (1994). We fit for a linear combination of the stellar continuum flux, F_{S99}, as

$$F^\star = \Sigma_{i=1}^{10} X_i\, F_i^{99}, \tag{7}$$

where X_i are the linear coefficients for a given age (i) and F_i^{99} are the STARBURST99 theoretical stellar continuum models for a given age.

Absorption lines of different ions are added using Voigt profiles defined by the: velocity shift v, b-parameter, column density N, and C_f. The metal covering fraction is initially fixed to 1 to reduce the number of free parameters. We include the O I absorption lines that are directly blended with the Lyman series. Each ion is considered independent, as are its parameters. For each galaxy, we test whether including the remaining metal lines listed in Table 2 improves the Lyman series fits. If, by eye, they do not improve the fits, we do not include the lines. Finally, the STARBURST99 models are convolved with the nominal spectral resolution of the observations.

We account for dust attenuation using the attenuation law from Reddy et al. (2016a), a uniform dust screen model, and fitting for the dust attenuation parameter ($E_{\rm B-V}$; similar to Chisholm et al. 2015). The linear combination of stellar continuum models, interstellar absorption lines, and dust attenuation produces the final fitted spectrum.

The data is fit using an IDL routine based on the non-linear least squares method, MPFIT (Markwardt 2009). MPFIT returns the best-fit and errors for $E_{\rm B-V}$, b, v, N and C_f of each ion. For observed spectra, the first step consists of masking the ISM absorption lines and the contaminating Milky Way absorption lines, geocoronal emission, and intervening absorbers. We apply these masks on the data and fit for the linear combination of dust-attenuated STARBURST99 models. In a second step, we fix the stellar continuum and fit for the ISM absorption lines as well as Milky Way absorption lines adjacent to the ISM lines. We fit for all of the observed Lyman series lines up to Ly6, provided that they are not near geocoronal emission or intervening absorbers, and do not have a signal to noise ratio below one. Since the simulated data do not contain the extra complications of adjacent Milky Way and geocoronal lines, all parameters are simultaneously fit in one step.

3.2. Fitting simulated data

We tested our fitting method with both noise-free and noisy synthetic data to determine how well the estimated parameter errors represent the actual errors. Additionally, we tested how these errors depend on the SNR and the spectral resolution. In Appendix A we fully describe the generation of synthetic spectra, but here we summarize the steps. The synthetic spectra were produced for two scenarios: one assuming the picket-fence model ($\log(N_{\rm H\,I}[{\rm cm}^{-2}]) = 20$, $C_f = 0.9$) and one describing a uniform ISM in the density-bounded regime ($\log(N_{\rm H\,I}[{\rm cm}^{-2}]) = 17.57$, $C_f = 1$). Both scenarios correspond to an escape fraction $f_{\rm esc} = 0.1$, regardless of the chosen dust distribution, since we set $E_{\rm B-V} = 0$ (Eqs. (5) and (6)).

We created synthetic spectra using the parameters of the picket-fence and density-bounded regime (see Table A.1) for seven different spectral resolutions between $R = 600 - 15000$. For each of these fourteen set-ups (7 spectral resolutions and two scenarios) we created 50 different realizations by adding different sets of random Gaussian noise. The level of Gaussian noise was chosen to produce a final SNR per pixel between 2 − 50, in seven total SNR steps. In total we created 98 different configurations (2 different scenarios, 7 spectral resolutions, and 7 SNRs), and each configuration has 50 individual spectra, for a total of 4900 synthetic spectra. Figs. B.1 and B.2 show synthetic spectra and the best-fit models for high and low spectral resolution ($R = 15000$ and $R = 600$) for three SNR configurations (noise-free, 10 and 2).

We fit the synthetic spectra with the methods outlined in Sect. 3.1.2. From this fitting we define two types of error:

- The *statistical error* ($Err_{\rm stat}$). This is the individual errors returned by MPFIT.
- The *systematic error* ($Err_{\rm syst}$). This is the deviation of the 50 parameter estimates of each scenario from the actual parameter value that created the original line profile. $Err_{\rm syst}$ is therefore a function of SNR and resolution. $Err_{\rm syst}$ is calculated as:

$$Err_{\rm syst} = \sqrt{\frac{1}{50}\sum_{i=1}^{50}(x_i - x_T)^2}, \tag{8}$$

where x_i is the estimated parameter and x_T is the true parameter value. The systematic error accounts for differences between the estimated parameters and the actual parameters that are unaccounted for by MPFIT. We express $Err_{\rm syst}$ as a percent error, defined as the deviation of the measured parameter from the true parameter divided by the true parameter value. This fractional error is more broadly applicable to a variety of measurement values.

By definition, $Err_{\rm syst}$ is not included in the statistical uncertainties reported by MPFIT, but it critically describes the ability of MPFIT to recover the actual parameter values. We account for $Err_{\rm syst}$ by defining the total error of each parameter (Err_{tot}) as the quadratic sum of $Err_{\rm syst}$ and $Err_{\rm stat}$:

$$Err_{\rm tot} = \sqrt{Err^2_{\rm syst} + Err^2_{\rm stat}}. \tag{9}$$

The systematic errors derived here are included in the errors of the H I covering fraction and column density (Cols. 3 and 6 of Table 3)

3.3. Fitting observed data

Using the fitting method described in Sect. 3.1.2, we fitted the observed spectra of the 18 galaxies in our sample. The resulting fit parameters $E_{\rm B-V}$, C_f, v, b, and N are listed in Table 3, and the corresponding best-fit spectra are shown in Appendix B. The reported uncertainties on the H I column densities and covering fractions account for the systematic errors derived from simulations.

We note that Tol0440-381 and Mrk54 have much lower redshifts than the rest of the sample. Consequently there are more Milky Way absorption lines near the Lyman series. This high-density of Milky Way Lyman series absorption lines means that we cannot accurately fit the Lyman series with Voigt profiles. Rather, we used non-parametric fits to describe the H I properties (see Sect. 3.3.1).

The H I column densities are affected by large uncertainties because the low SNR and insufficient resolution of the spectra does not allow a reliable estimate of $N_{\rm H\,I}$ given the saturation of the H I absorption lines (see Sect. 4.1). Therefore direct $N_{\rm H\,I}$ measurements from the Lyman series are not reliable, and an indirect method is used (Sect. 3.3.2). To examine whether saturation and the resulting degeneracies affect the C_f or metal column densities that we derived, we refit the spectra fixing the H I column density to $N_{\rm H\,I}$ = 10^{18} cm^{-2}. This $N_{\rm H\,I}$ corresponds to an optically thick portion of the curve of growth for the observed Lyman series lines (from Lyβ to Ly6). We find that all the fit parameters (except for b(H I)) are consistent within 1σ with the results listed in Table 3. The fact that b changes is consistent with b being degenerate with $N_{\rm H\,I}$ at these column densities. This indicates that saturated H I absorption does not affect the measured velocities and covering fractions.

3.3.1. Non-parametric measurements of the H I covering fraction

Fitting the Lyman series using the the radiative transfer equation (Eq. (1)) is one way to measure the H I covering fraction. However, this assumes that the absorption profiles follow a single Lorentzian velocity distribution. Observed absorption profiles arising from galactic outflows are highly non-Lorentzian (Heckman et al. 2000; Pettini et al. 2002; Shapley et al. 2003; Weiner et al. 2009; Chisholm et al. 2017b). Consequently, we also measure C_f from the residual flux of the Lyman series lines after removing the stellar continuum (see Eq. (3)). This assumes a uniform dust geometry and that the Lyman series is fully saturated. The Lyman series is saturated from $N_{\rm H\,I} \gtrsim 10^{16}$ cm^{-2} (for Lyβ to Ly6). Importantly, this non-parametric approach does not assume a velocity distribution of the H I gas, and accounts for arbitrary line profiles.

The covering fraction is derived as the maximum of $(1-F_{\rm Gau})$ in a velocity range chosen by eye near the deepest part of each Lyman series line. $F_{\rm Gau}$ is the stellar continuum removed flux, modified by a Gaussian kernel centered on zero with standard deviation corresponding to the error array. We measure $(1-F_{\rm Gau})$ 1000 times, where each time the flux value of each pixel is determined from a different noise distribution. We then take the median and the standard deviation of this distribution as the C_f value and uncertainty for each Lyman series transition. Table C.3 lists the C_f derived from the residual flux of each Lyman series transition in each galaxy. We then define the C_f(H I) from the residual flux as the error weighted mean (M_W) of the i observed Lyman series transitions and the C_f error as the error on M_W as:

$$M_W = \frac{\sum_{i=1}^{n} C_{f_i} \times \omega_i}{\sum_{i=1}^{n}\omega_i} \text{ with } \omega_i = \frac{1}{\sigma_i^2}, \qquad \sigma_{M_W} = \sqrt{\frac{1}{\sum_{i=1}^{n}\omega_i}}. \tag{10}$$

The corresponding values, denoted as "C_f depth", are reported in Table 3. The C_f depth values are consistent, within the errors, with the C_f values derived from the fits in Sect. 3.3. Except for Tol0440-381 and Mrk54, where we do not fit the absorption lines with the method in Sect. 3.1.2, we compute the final C_f(H I) as the weighted mean of the C_f depth and the fitted C_f values (column 8 in Table 3).

3.3.2. Indirect measurements of the H I column density

Since the Lyman series lines are saturated, but not damped, and the spectra have insufficient spectral resolution and SNR (cf. Sect. 4.1), direct H I column density measurements are largely unconstrained. Therefore, indirect methods of measuring $N_{\rm H\,I}$ are needed. We use the O I absorption lines, constrained when possible by the unsaturated O I 1039Å absorption line, to derive the O I column density. Using the known metallicity (12+log(O/H)) of the galaxy, we then indirectly infer the hydrogen column density. This approach assumes that the emission-line based oxygen abundance, tracing the chemical composition of the ionized gas, is identical to that of the neutral gas. If the metallicity was lower in the neutral gas, e.g. due to the presence of some pristine gas

Table 3: Derived H I properties from the Lyman series absorption lines.

Galaxy name	$E_{\mathrm{B-V}}$	$\log(N_{\mathrm{H\,I}})$ [log(cm^{-2})]	b [km s^{-1}]	v [km s^{-1}]	C_f fits	C_f depth	C_f final
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
J0921+4509	0.222 ± 0.015	17.94 ± 3.22	95 ± 49	-56 ± 13	0.754 ± 0.111	0.769 ± 0.116	0.761 ± 0.080
J1503+3644	0.274 ± 0.014	20.50 ± 2.33	91 ± 24	-127 ± 12	0.634 ± 0.109	0.754 ± 0.062	0.725 ± 0.054
J0925+1409	0.164 ± 0.015	17.81 ± 3.26	81 ± 73	-214 ± 151	0.652 ± 0.218	0.635 ± 0.094	0.638 ± 0.086
J1152+3400	0.134 ± 0.022	20.94 ± 2.70	187 ± 48	-468 ± 39	0.506 ± 0.067	0.619 ± 0.08	0.548 ± 0.053
J1333+6246	0.151 ± 0.043	21.14 ± 1.46	102 ± 19	-126 ± 48	0.731 ± 0.122	0.826 ± 0.066	0.804 ± 0.058
J1442-0209	0.140 ± 0.015	19.60 ± 3.53	123 ± 183	-261 ± 34	0.621 ± 0.120	0.549 ± 0.040	0.556 ± 0.038
Tol1247-232	0.156 ± 0.010	21.89 ± 3.09	66 ± 87	260 ± 31	0.587 ± 0.061	0.690 ± 0.084	0.623 ± 0.049
Tol0440-381	0.271 ± 0.028	-	-	-	-	0.570 ± 0.084	0.570 ± 0.084
Mrk54	0.359 ± 0.013	-	-	-	-	0.504 ± 0.077	0.504 ± 0.077
J0926+4427	0.114 ± 0.010	16.39 ± 0.40	213 ± 23	-199 ± 12	0.723 ± 0.048	0.814 ± 0.048	0.768 ± 0.034
J1429+0643	0.108 ± 0.015	16.17 ± 0.50	236 ± 56	-241 ± 36	0.897 ± 0.071	0.960 ± 0.061	0.930 ± 0.046
GP0303-0759	0.121 ± 0.045	16.07 ± 1.50	192 ± 99	-266 ±92	0.908 ± 0.207	-	0.908 ± 0.207
GP1244+0216	0.290 ± 0.043	16.34 ± 0.80	220 ± 79	-78 ± 48	0.909 ± 0.357	0.950 ± 0.131	0.946 ± 0.123
GP1054+5238	0.204 ± 0.044	19.60 ± 3.81	164 ± 78	-166 ± 29	0.702 ± 0.131	0.889 ± 0.158	0.778 ± 0.101
GP0911+1831	0.352 ± 0.038	16.76 ± 1.19	188 ± 43	-273 ± 40	0.731 ± 0.150	0.765 ± 0.116	0.752 ± 0.092
SGAS J1226	0.201 ± 0.001	17.16 ± 1.06	380 ± 24	-264 ± 21	0.932 ± 0.038	0.998 ± 0.009	0.994 ± 0.009
SGAS J1527	0.370 ± 0.002	17.15 ± 1.22	269 ± 47	-208 ± 25	1.000 ± 0.075	0.990 ± 0.038	0.992 ± 0.034
Cosmic Eye	0.405 ± 0.001	22.59 ± 1.28	199 ± 5	56 ± 16	0.918 ± 0.072	0.998 ± 0.024	0.990 ± 0.023

Notes. (1) Galaxy name; (2) dust attenuation parameter ($E_{\mathrm{B-V}}$); (3) logarithm of the H I column density; (4) H I Doppler b-parameter; (5) H I velocity shift; and (6) H I covering fraction from the fits. Uncertainties of the H I column density and covering fraction from the fits include the systematic errors detailed in Sect 4.2. (7) H I covering fraction measurements derived from the residual flux of the individual H I absorption lines (Sect. 3.3.1; Table C.3). GP0303-0759 does not have a reliable depth measurement because Milky Way absorption lines overlap the Lyman series; (8) weighted mean of (6) and (7).

(see, e.g. Lebouteiller et al. 2013, and references therein) or incomplete mixing of the metals (see Sect. 5.2), the resulting $N_{\mathrm{H\,I}}$ would be *larger* (and more saturated) than inferred here.

The O I fit parameters b, v, $N_{\mathrm{O\,I}}$, and the H I column density derived using the metallicities from Table 1, are listed in Table 4. A curve of growth analysis indicates that O I 1039Å saturates at $N_{\mathrm{O\,I}} > 10^{16.5}$ cm^{-2}, whereas all but one of our $N_{\mathrm{O\,I}}$ are less than $10^{16.2}$ cm^{-2}. If the O I 1039 line is saturated, then the $N_{\mathrm{H\,I}}$ values in Table 4 would under-estimate the actual $N_{\mathrm{H\,I}}$ (i.e. $N_{\mathrm{H\,I}}$ is not lower than the quoted values). O I is undetected for J0925+1409 (see Fig. B.5), and we use O I 1302 Å for Tol0440-381 and Mrk54. We could not determine $N_{\mathrm{H\,I}}$ for the three MegaSaura sources, since the O I line was either saturated, or there was not a literature metallicity. The resulting H I column densities are found to be $\log(N_{\mathrm{H\,I}}) \sim 18.6 - 20$ cm^{-2}.

We fixed the O I covering fractions to the H I values (i.e. $C_f(\mathrm{O\,I}) = C_f(\mathrm{H\,I})$). This is plausible because O I and H I have similar ionization potentials and their ionization fractions are locked together by charge exchange. We examined how tying the covering fractions impact the derived O I column densities. As an extreme case, we fixed the O I covering fraction to 1 for J1152+3400, which has one of the lowest $C_f(\mathrm{H\,I})$, and J0921+4509, which has the lowest $N_{\mathrm{H\,I}}$ derived from O I. We found that $\log(N_{\mathrm{OI}})$ is reduced by 0.30 and 0.15 dex, respectively, which is comparable to the $N_{\mathrm{O\,I}}$ errors. Therefore, we conclude that tying $C_f(\mathrm{O\,I})$ to $C_f(\mathrm{H\,I})$ does not drastically change the measured O I column density.

3.3.3. The Si II covering fraction

We measured the Si II covering fraction ($C_f(\mathrm{Si\,II})$) using the Si II 1190 Å doublet and the Si II 1260 Å singlet. For the Si II 1260 Å singlet, $C_f(\mathrm{Si\,II})$ is derived using the same procedure as for the residual flux measurements of the Lyman series (Sect. 3.3.1). A different approach, detailed in Chisholm et al. (2017b), is adapted for the doublet. As the two transitions share the same C_f and the ratio of the two optical depths is given by the ratio of their oscillator strengths, f, the velocity-resolved covering fraction is measured from a system of equations (Hamann et al. 1997) as:

$$C_f(v) = \frac{F_W(v)^2 - 2F_W(v) + 1}{F_S(v) - 2F_W(v) + 1} \tag{11}$$

where F_W is the continuum subtracted flux of the weaker doublet line (Si II 1190 Å) and F_S is the continuum subtracted flux of the stronger doublet line (Si II 1193 Å). This method accounts for the possibility that the lines are not saturated. We measure the covering fraction 1000 times by varying the flux in a similar way to the residual flux measurements of the Lyman series. The median and standard deviation of this distribution is taken as the $C_f(\mathrm{Si\,II})$ value and error (Table 5).

3.4. Effect of the UV attenuation law on the covering fraction

A priori the measurements made here also depend on the dust attenuation law used. In this study, we used the attenuation law from Reddy et al. (2016a) because it is defined blueward of Lyα.

Table 4: Fitted O I properties

Galaxy name (1)	$\log(N_{\mathrm{O\,I}})$ [log(cm^{-2})] (2)	b [km s^{-1}] (3)	v [km s^{-1}] (4)	$\log(N_{\mathrm{H\,I}})$ [log(cm^{-2})] (5)
J0921+4509	15.30 ± 0.13	45 ± 15	62 ± 11	18.63 ± 0.19
J1503+3644	15.55 ± 0.16	302 ± 98	102 ± 77	19.60 ± 0.17
J0925+1409	-	-	-	-
J1152+3400	15.43 ± 0.17	227 ± 129	-102 ± 83	19.43 ± 0.18
J1333+6246	15.54 ± 0.35	287 ± 277	-213 ± 152	19.78 ± 0.37
J1442-0209	15.62 ± 0.58	178 ± 145	82 ± 101	19.69 ± 0.58
Tol1247-232	15.29 ± 0.43	278 ± 168	58 ± 143	19.19 ± 0.44
Tol0440-381[a]	15.47 ± 0.02	623 ± 28	12 ± 19	19.27 ± 0.10
Mrk54[a]	16.07 ± 0.01	619 ± 8	15 ± 6	19.37 ± 0.10
J0926+4427	15.77 ± 0.02	118 ± 5	-141 ± 4	19.76 ± 0.05
J1429+0643	15.55 ± 0.24	218 ± 74	-43 ± 98	19.35 ± 0.25
GP0303-0759	15.41 ± 0.18	209 ± 84	31 ± 70	19.55 ± 0.19
GP1244+0216	16.12 ± 0.14	157 ± 47	-67 ± 36	19.95 ± 0.15
GP1054+5238	15.73 ± 0.14	256 ± 72	39 ± 55	19.63 ± 0.15
GP0911+1831	15.73 ± 0.15	221 ± 79	-102 ± 67	19.73 ± 0.16
SGAS J1226	16.11 ± 0.03	200 ± 7	-290 ± 11	-
SGAS J1527	15.57 ± 0.13	133 ± 44	-89 ± 33	-
Cosmic Eye	20.64[b] ± 0.66	24 ± 77	208 ± 9	24.04[b] ± 0.67

Notes. (1) Galaxy name; (2) logarithm of O I column density; (3) O I Doppler b-parameter; (4) velocity shift of the O I line. The O I covering fraction is fixed to the H I final value from Table 3. We did not detect the O I absorption lines in the J0925+1409 spectra. (5) H I column density derived from the product of (3) and 12 + log(O/H) (Table 1). Note that we cannot estimate $N_{\mathrm{H\,I}}$ for SGAS J1226 and SGAS J1527 because they do not have a measured 12 + log(O/H).
[(a)] Used O I 1302 Å
[(b)] Saturated O I line, hence unreliable $N_{\mathrm{H\,I}}$ determination. From the damped Lyα profile, Quider et al. (2010) obtained $N_{\mathrm{H\,I}} = (3.0 \pm 0.8) \times 10^{21}$ cm^{-2}.

We refit J1503+3644 and GP0911+1831, two galaxies with high $E_{\mathrm{B-V}}$ and low C_f, using a Small Magellanic Cloud (SMC) attenuation law[1], still assuming a uniform dust foreground. The SMC law is significantly steeper than the Reddy et al. (2016a) law. With the SMC dust law, we measure C_f(H I) of 0.653 ± 0.109 and 0.744 ± 0.156, respectively. These are consistent, within 1 σ, with the C_f(H I) estimated using the Reddy et al. (2016a) attenuation law (0.634 ± 0.109 and 0.731 ± 0.150). The $E_{\mathrm{B-V}}$ values will change based upon the attenuation law used to match the observed continuum, but these changes do not affect the measured C_f. We therefore conclude that the adopted attenuation law does not significantly impact the measured covering fractions.

4. Recovering H I properties from simulated spectra

We simulated synthetic spectra and fit these mock H I lines with the method in Sect. 3.2. Comparing the fitted results with the parameters that created the spectra characterizes how accurately the method returns the H I parameters. Here, we discuss these results in the context of the H I column density (Sect. 4.1) and the covering fraction (Sect. 4.2). This discussion illustrates that C_f is accurately measured for most resolutions and SNRs, while the H I column density has large uncertainties. These simulations are especially helpful for planning future observations by determining the SNRs and resolutions required to accurately measure the covering fractions of LyC emitters.

[1] Values have been taken from the IDL routine from J. Xavier Prochaska: `https://github.com/profxj/xidl/tree/master/Dust`

4.1. H I *column densities*

The synthetic spectra allow us to quantify how accurately our method reproduces the H I properties. For SNRs (per pixel) less than 10 and resolutions less than 3000, the simulations have $N_{\mathrm{H\,I}}$ percent errors greater than 300% (Table C.1). Even at higher resolutions (R = 15000), the percent error is greater than 200%, unless the SNR is greater than 5. For the lower SNRs typical of our observations, we measure order of magnitude systematic uncertainties on $N_{\mathrm{H\,I}}$. These large uncertainties are inherent because the Lyman series transitions saturate for H I column densities between $N = 10^{16}$ cm^{-2} to $N = 10^{22}$cm^{-2}. For these so-called Lyman limit systems, high quality and very high resolution spectra ($R \sim 30000$) are needed to constrain $N_{\mathrm{H\,I}}$ with Voigt fitting methods (see e.g. O'Meara et al. 2007). Therefore, we conclude that $N_{\mathrm{H\,I}}$ cannot be directly fitted from the Lyman series absorption lines. However, the O I absorption lines included in our fits remain unsaturated for $N_{\mathrm{O\,I}} < 10^{16.5}$ cm^{-2}, and do not suffer from these large uncertainties. Consequently, the neutral column density is most accurately inferred by converting $N_{\mathrm{O\,I}}$ into $N_{\mathrm{H\,I}}$ using the gas-phase metallicity, as done in Sect. 3.3.2.

Table 5: Si II covering fractions

Galaxy name (1)	Si II 1190 Å (2)	Si II 1260 Å (3)	Mean (4)
J0921+4509	0.66 ± 0.32	0.59 ± 0.15	0.60 ± 0.14
J1503+3644	0.38 ± 0.35	0.56 ± 0.45	0.45 ± 0.28
J0925+1409	0.40 ± 0.28	0.43 ± 0.34	0.41 ± 0.19
J1152+3400	0.27 ± 0.26	0.23 ± 0.55	0.27 ± 0.24
J1333+6246	0.29 ± 0.26	0.56 ± 0.35	0.39 ± 0.21
J1442-0209	0.46 ± 0.23	0.48 ± 0.34	0.47 ± 0.19
Tol1247-232	-	0.26 ± 0.01	0.26 ± 0.01
Tol0440-381	-	0.37 ± 0.04	0.37 ± 0.04
Mrk54	-	0.32 ± 0.01	0.32 ± 0.01
J0926+4427	0.37 ± 0.19	0.36 ± 0.06	0.36 ± 0.06
J1429+0643	0.73 ± 0.26	0.78 ± 0.11	0.77 ± 0.10
GP0303-0759	0.54 ± 0.29	0.47 ± 0.12	0.48 ± 0.11
GP1244+0216	0.49 ± 0.39	0.51 ± 0.18	0.50 ± 0.16
GP1054+5238	0.41 ± 0.34	0.49 ± 0.22	0.46 ± 0.19
GP0911+1831	0.39 ± 0.36	0.45 ± 0.33	0.43 ± 0.24
SGAS J1226	0.97 ± 0.06	0.91 ± 0.04	0.93 ± 0.03
SGAS J1527	0.83 ± 0.16	0.79 ± 0.10	0.80 ± 0.08
Cosmic Eye	0.98 ± 0.06	0.94 ± 0.03	0.95 ± 0.02

Notes. (1) Galaxy name; (2) C_f(Si II) derived from the Si II 1190 Å doublet using Eq. (11); (3) C_f(Si II) derived from the Si II 1260 Å absorption line; (4) weighted mean between (2) and (3). We do not observe the Si II 1190 Å doublet for Tol1247-232, Tol0440-381 and Mrk54 because these lines fall in the COS detector gap.

4.2. H I covering fractions

Conversely, the simulations show that C_f(H I) has a low percent error, under typical observing conditions. At $R > 1500$ and SNR > 5, the C_f systematic percent errors are less than 6% of the measured value (Fig. 2, Table C.2). Therefore, the neutral gas covering fractions are accurately recovered from our observational conditions.

JWST will accurately measure C_f of metal absorption lines from high-redshift galaxies. NIRSpec on *JWST* is expected to have $R \sim 3000$ (1000) in the high (medium) resolution configurations. This means that the systematic errors will be 3% (6%) of the measured C_f for SNR = 5 observations, illustrating the feasibility of measuring C_f from high-redshift galaxies with *JWST*.

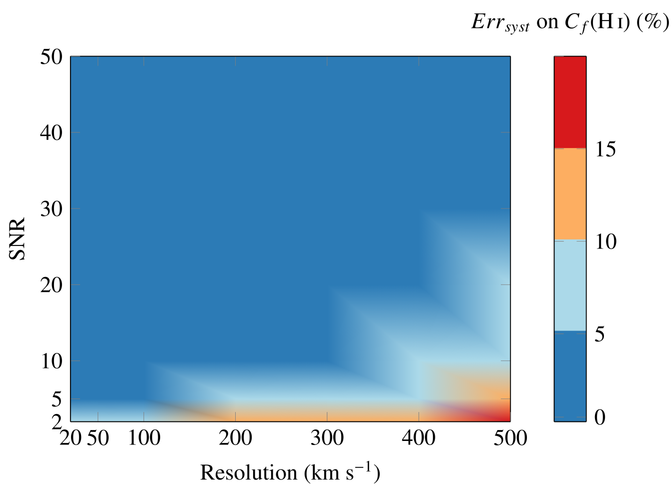

Fig. 2: Color map of the systematic percent error of the covering fraction as a function of the resolution and signal-to-noise ratio (SNR). The synthetic spectra are created with R < 120 km s^{-1} combine a theoretical stellar continuum spectra with R(S99) = 120 km s^{-1} and absorption lines of spectral resolution R. The covering fraction is recovered to within 5% of the estimated parameter for all observations within the dark blue region.

5. The covering fraction of LyC emitters and comparison sources

We now examine the derived covering fractions from the Lyman series and the Si II absorption lines. Then we discuss different geometrical model assumptions and compare our results to earlier work.

5.1. Leakers have low neutral gas covering fractions

The H I covering fraction describes the porosity of the neutral gas, and demonstrates whether the neutral gas is clumpy. A smaller C_f(H I) means that there are more low-density channels for ionizing photons to escape through.

Our sample has H I covering fractions ranging from 0.50 to unity (Fig. 3). Only five of the 18 galaxies (28%) have an H I covering fraction consistent with unity at 1σ. The low C_f(H I) values are likely because the sample is biased: 15 of these 18 galaxies were targeted as potential LyC leaker, or for being particularly strong line emitters. Since a non-unity C_f is a possible LyC escape mechanism, it is not too surprising that many of these galaxies have low C_f(H I). The three galaxies that were not targeted as LyC leakers (the MegaSaura galaxies) have C_f(H I) consistent

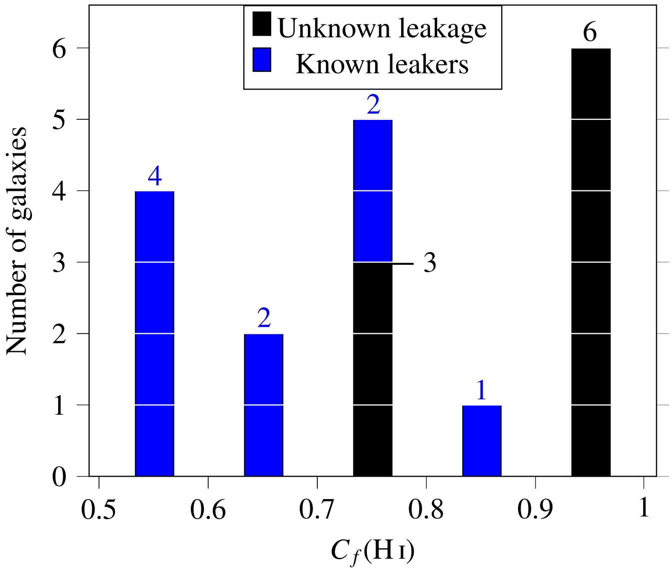

Fig. 3: Histogram of the H I covering fraction (C_f(H I)) from the 18 galaxies in our sample. We split the sample into galaxies known to leak LyC photons (blue) and galaxies without measured LyC emission (black). The leakers have the lowest C_f(H I) and the unknown emitters have the highest C_f(H I).

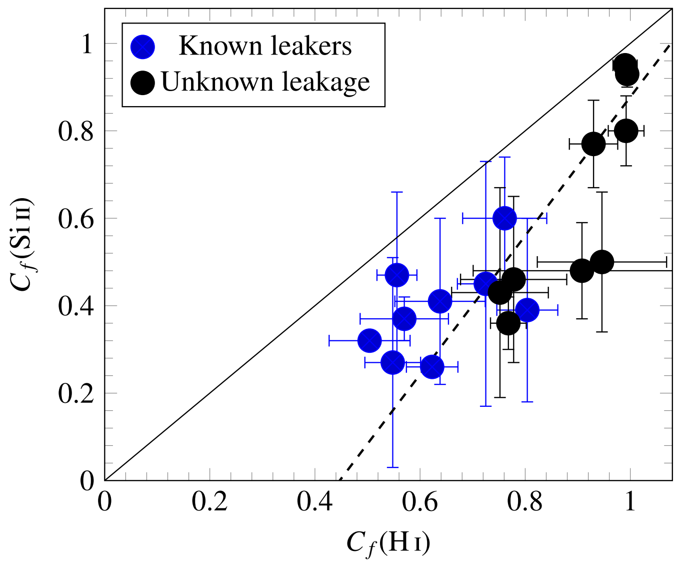

Fig. 4: Comparison of the Si II and H I covering fractions. The black solid line shows a one-to-one relation and the dotted line is the fitted relation (Eq. (12)). The average offset between C_f(H I) and C_f(Si II) is 0.25.

with unity. In contrast, the galaxies with the largest confirmed escape fractions of ionizing photons have the lowest C_f(H I) values (Fig. 3). The leakers have a median C_f(H I) = 0.62 ± 0.10, while the unknown leakers have a median C_f(H I) = 0.95 ± 0.10. Lyman continuum emitters have low H I covering fractions, which allows LyC photons to escape through low density channels.

The large $N_{\rm H\,I}$ values, estimated from the O I column densities (Table 4), further emphasizes that ionizing photons must escape through holes in the H I. $N_{\rm H\,I}$, calculated from $N_{\rm O\,I}$, is greater than $10^{18.6}$ cm^{-2} for the entire sample. Even without converting into $N_{\rm H\,I}$, the large $N_{\rm O\,I}$ values require unphysically large metallicities (12+log(O/H) > 10) for ionizing photons to escape through low density regions. At these column densities the Lyman series and Lyman limit are saturated. Ionizing photons cannot pass through the neutral gas unabsorbed. In other words, even LyC galaxies have optically thick H I; ionizing photons must escape through low-density channels.

Low C_f(H I) values indicate that the escape of LyC photons is dominated by a patchy ISM, or the Picket-Fence model. However, we find that a low covering fraction is not the only parameter leading to a large $f_{\rm esc}$. For example, Mrk 54 has a C_f(H I) ~0.5 and a $f_{\rm esc}$ < 1%. Dust crucially impacts the LyC escape fraction by removing ionizing photons (Eq. (5)). Consequently, the escape fractions cannot simply be inferred from the measured covering fractions. This is discussed in depth in Sect. 5.3 and Paper II.

Using Eq. (5), we can predict which of the 9 galaxies in the sample without measured LyC emission should emit ionizing photons. To emit LyC photons, both a low $E_{\rm B-V}$ and a low C_f are required. J0926+4427, J1429+0643, and GP1054+5238 are the best candidates in our sample to leak ionizing photons (in order of most likely to emit ionizing photons). Follow-up observations should expect to find $f_{\rm esc}$ values between 0.02 – 0.06.

5.2. H I and Si II covering fractions

Low-ionization metal absorption lines are often used as proxies for the neutral gas covering fraction (Shapley et al. 2003; Jones et al. 2013; Alexandroff et al. 2015; Trainor et al. 2015). Metal lines are in redder portions of the spectra. Consequently, instruments can more efficiently observe metals lines at low-redshifts, and the Lyman forest does not obscure metal lines at high-redshifts. Therefore, metal absorption lines are ideal probes of neutral gas properties. However, recent observations of $z \sim 3$ stacked spectra suggest that metal absorption lines have covering fractions a factor of two smaller than H I absorption lines (Reddy et al. 2016b). Here, we test whether C_f(Si II) traces C_f(H I).

The C_f(Si II) is systematically lower than C_f(H I), with a mean offset of 0.25 (Fig. 4). C_f(H I) and C_f(Si II) are linearly related at the 3σ significance level (p-value < 0.001; Pearson's correlation coefficient of 0.79) as:

$$C_f(\text{H I}) = (0.63 \pm 0.12) \times C_f(\text{Si II}) + (0.44 \pm 0.07). \quad (12)$$

While C_f(Si II) is not equal to C_f(H I), Eq. (12) estimates C_f(H I) from C_f(Si II). Neutral gas covering fractions can be estimated from metal covering fractions.

There are numerous reasons why C_f(Si II) could be related to, but not equal to, C_f(H I) (see Reddy et al. 2016b, for an in-depth discussion). First, narrow, high covering fraction, metal lines may be unresolved in low-resolution spectra (~1500). While this is possible for low-resolution data, it is less likely for the high-resolution COS data. The Si II lines could possibly be unsaturated, but the doublet method accounts for this possibility and does not remove the systematic offset. Alternatively, the Si II ionization potential overlaps with, but is not equal to, H I. Therefore, the Si II and H I gas may trace similar, but not equal, gas. However, Reddy et al. (2016b) find that the Si II and H I line profiles and central velocities are similar (also see Chisholm et al. 2016). This indicates that the two transitions are co-moving and trace similar gas. Finally, the Si II lines are also potentially affected by scattering and fluorescent emission in-filling, although it is difficult to predict the amount the emission increases the covering fraction (cf. Prochaska et al. 2011; Scarlata & Panagia 2015).

Another explanation, is that the neutral gas and metal-enriched gas are not fully mixed. In this situation, Si II only probes metal-rich regions, while the H I gas probes both high and low metallicity gas. Therefore, the neutral gas covers more of the background source than the metals alone. At lower metallicities this is exaggerated because there is less metal-rich gas. This leads to fewer high-density metal regions to absorb the background continuum, and systematically reduces the metal covering fraction. To test the effect of metallicity on the relation between the neutral and metal covering fractions, we fit a multiple linear relationship between C_f(H I), C_f(Si II), and 12 + log(O/H). We find a significant (3σ, p-value < 0.001, Pearson's correlation coefficient of 0.79) trend between these three parameters that scales as:

$$C_f(\text{H I}) = (1.8 \pm 0.8) - (0.18 \pm 0.10) \times [12 + \log(\text{O/H})] + (0.75 \pm 0.16) \times C_f(\text{Si II}) \quad (13)$$

While Eq. 13 is only marginally more significant than the simpler relation between C_f(Si II) and C_f(H I), it physically explains the relation between the neutral and metal covering fractions.

The difference between Eq. (12) and Eq. (13) is the metallicity dependence of the correction added to C_f(Si II). As the metallicity increases, the required correction to convert C_f(Si II) into C_f(H I) decreases. While at lower metallicities, the factor added to C_f(Si II) must be larger. Physically, this means that at lower metallicities the Si II traces a smaller fraction of the total H I. Therefore, the linear relationship between C_f(H I) and C_f(Si II) may arise because metal-rich clumps cover the background source differently than the H I gas.

The empirical relationship, Eq. (12), is recovered from Eq. (13) when the median metallicity of the sample (8.06) is used. However, Eq. (12) would change if the median 12 + log(O/H) of the sample changes. This suggests that the empirical relationship between C_f(H I) and C_f(Si II) is only constrained near the metallicities of the sample. For divergent 12 + log(O/H) values, the metallicity should be accounted for when estimating C_f(H I).

5.3. Is the dust geometry clumpy or uniform?

In Sect. 5.1, we emphasized that dust heavily contributes to the escape of ionizing photons. However, since the impact of dust depends on the assumed geometry, we explored the two cases described above (Sect. 3.1.1), namely *(a)* a uniform dust screen and *(b)* a clumpy geometry dust model. To test the impact of these assumed geometries on C_f and $E_{\text{B-V}}$, we modified our fitting routine from *(a)* a uniform dust screen model to *(b)* a clumpy model. We refit two galaxies, J0921+4509 and J1152+3400, one galaxy with an average C_f(H I) and one with a low C_f(H I). Adopting a clumpy geometry *(b)*, we derive $E_{\text{B-V}}$= 0.236 and 0.239 (versus $E_{\text{B-V}}$ = 0.224 and 0.134 in model *(a)*) and C_f(H I) = 0.976 and 0.912 (versus C_f = 0.754 and 0.506 in model *(a)*), respectively. These values and Eq. (6) predict a LyC escape fraction of f_{esc} = 0.024 and 0.088 at λ = 912 Å, consistent, within 1σ, with the values derived using the uniform dust screen model ($f_{\text{esc}} = 0.017 \pm 0.007$% and 0.092 ± 0.027%).

As just shown, using the uniform dust screen leads to a lower $E_{\text{B-V}}$ and C_f(H I) compared to the clumpy model. In the clumpy model, all of the light that escapes through holes in the neutral gas is unattenuated. Therefore, the holes in the neutral gas must be smaller, while any continuum passing through the neutral gas must be more heavily attenuated to match the observed flux. C_f(H I) and $E_{\text{B-V}}$ must take values that fit both the geometry and the observed flux (extending to Lyman continuum). Both quantities are therefore model dependent. However, disentangling the actual geometric model is not mandatory to predict f_{esc} with C_f(H I) and $E_{\text{B-V}}$, as long as the geometry is consistently modeled (see Paper II).

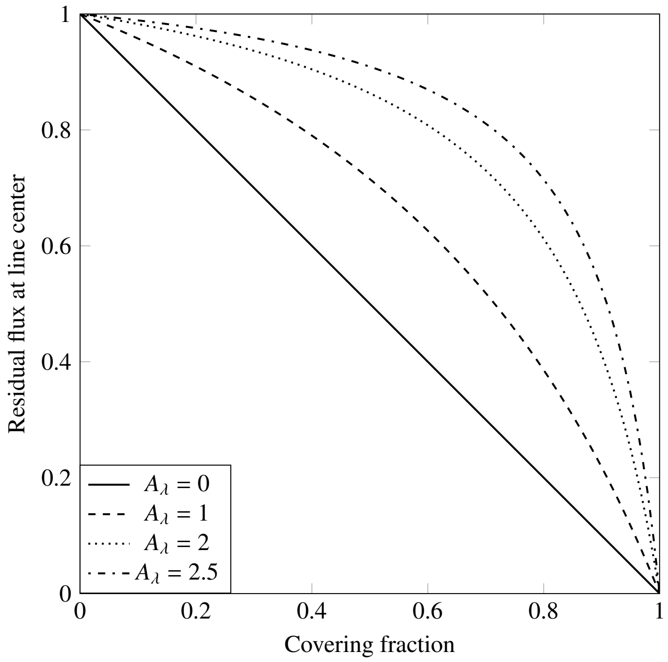

Fig. 5: Residual flux R of a saturated absorption line from Eq. (4) for different values of $A_\lambda = k_\lambda E_{\text{B-V}}$ in a clumpy ISM geometry. When dust is present, the residual flux is larger than $(1 - C_f)$. However, in a uniform dust screen geometry, the residual flux is always $R = (1 - C_f)$. The covering fraction derived from the observed residual flux and a clumpy dust geometry is underestimated when dust is present.

On the other hand, the observed residual flux of the H I absorption lines provide clues to the geometric model. Eq (4) shows that the residual flux of the H I absorption lines are wavelength dependent in the clumpy geometry. Therefore, if the residual flux of the individual Lyman series absorption lines change with wavelength (for example from Lyβ to Ly6), then the data would suggest a clumpy dust geometry. Fig. 5 shows the relationship between the residual flux of a saturated absorption line (R) and the neutral covering fraction for different values of $A_\lambda = k_\lambda E_{\text{B-V}}$ in a clumpy geometry. As A_λ increases (larger $E_{\text{B-V}}$ or a bluer transition), the residual flux decreases as dust removes more continuum photons. In the clumpy dust geometry, R is less than 1-C_f if there is dust. For a covering fraction of 0.8, the measured R is 0.2 when A_λ is 0, 0.4 when A_λ is 0.1, and 0.7 when A_λ is 0.2 (Fig. 5). The C_f estimated from 1-R is underestimated in the clumpy dust geometry case.

Since the residual intensity increases with A_λ and k_λ, this effect, if present, should be detectable in the Lyman series absorption lines. Nevertheless, the differences between the residual flux of the Lyman series should be small, and most easily detected in galaxies with large $E_{\text{B-V}}$. Table 6 shows the expected residual flux of saturated Lyman series lines, assuming an $f_{\text{esc}} = 0.10$ (C_f = 0.9 in the clumpy geometry). When $E_{\text{B-V}}$ is 0.1, the differ-

Table 6: Theoretical residual flux for the Lyman series (Lyβ to Ly6) of a galaxy with a clumpy dust geometry.

H I lines	$E_{\rm B-V}$			
	0.0	0.1	0.2	0.3
Lyβ	0.100	0.245	0.489	0.737
Lyγ	0.100	0.254	0.512	0.764
Lyδ	0.100	0.259	0.523	0.775
Ly5	0.100	0.261	0.530	0.782
Ly6	0.100	0.262	0.533	0.785

Notes. Theoretical measurements of the residual flux of a saturated absorption line for five Lyman series lines. These values are calculated assuming $f_{\rm esc}$ = 0.10, a clumpy dust geometry, saturated Lyman lines, and the Reddy et al. (2016a) dust attenuation law. The differences found between different Lyman lines for a given $E_{\rm B-V}$ are small and accurate residual flux measurements are required to distinguish the geometries.

ence between the Lyβ and Ly6 residual flux is less than 0.02. For $E_{\rm B-V}$= 0.30, the offset rises to 0.05. While we do not measure a trend in our observations (Table C.3), the measured residual intensities have large uncertainties. Our simulations imply that SNR = 30 and R = 15000 are required to distinguish the two geometries. We cannot determine the dust geometry with the current data, and both a uniform dust screen or clumpy geometry are allowed by the observations.

5.4. Comparison with other studies

Reddy et al. (2016b) recently determined the H I covering fraction of $z \sim 3$ Lyman break galaxies from Lyman series fits. Analyzing stacked spectra, they found high C_f(H I) ∼ 0.92 – 0.97. They also examined the relation between C_f(H I) and the C_f of metal lines (Si II, C II, and Al III), finding a much lower covering fraction for the metals than for H I. The differences between their C_f(metal) and their C_f(H I) are nearly twice as large as we find for C_f(Si II) in Fig. 4 (their C_f(metal) is 0.4-0.6 smaller than C_f(H I), versus our average of 0.25).

This discrepancy arises from an inconsistency in determining C_f(metal). To fit their C_f(H I), Reddy et al. (2016b) adopt a clumpy geometry with dust only in the H I. The continuum level used to obtain the normalized composite spectra in their study is determined as:

$$F_\lambda^{\rm cont} = F_\lambda^\star 10^{-0.4E_{\rm B-V}k_\lambda} \times C_f({\rm H\,I}) + F_\lambda^\star[1 - C_f({\rm H\,I})] \quad (14)$$

The residual flux of a particular metal absorption line is

$$R({\rm metal}) = \frac{F_\lambda}{F_\lambda^{\rm cont}} = \frac{1 - C_f({\rm metal})}{10^{-0.4E_{\rm B-V}k_\lambda} \times C_f({\rm H\,I}) + 1 - C_f({\rm H\,I})} \quad (15)$$

Note that Eq. (15) reduces to Eq. (4) when C_f(metal) = C_f(H I). In this geometry, the residual flux of a saturated absorption lines is *not* simply 1-R. The covering fraction cannot be read off from the profile; a measured $E_{\rm B-V}$ is required to determine C_f (see the large differences in Fig. 5; Vasei et al. 2016). However, Reddy et al. (2016b) measure C_f(metal) as $1 - R$ from the continuum normalized spectra. Meanwhile, they determine C_f(H I) from the full fits of their spectra, which properly accounts for their assumed geometry. This means that C_f(metal) and C_f(H I) are calculated with different geometric assumptions.

The inconsistency in the derivation of C_f(metal) and C_f(H I) produces the large mismatch between our observations and

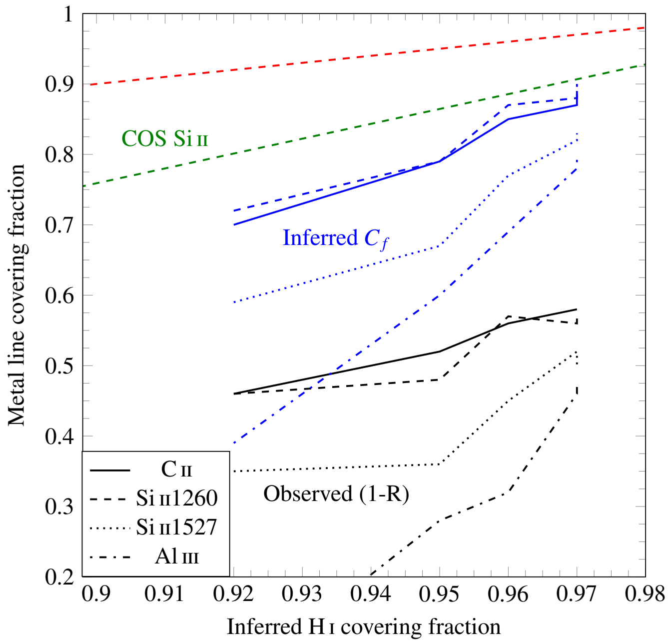

Fig. 6: Metal covering fractions from Si II, C II and Al III (as given by the line-types in the legend) as a function of the H I covering fraction (C_f(H I)) from Reddy et al. (2016b). C_f(H I) is always measured in a clumpy geometry. The black lines show $1-R$ measured from the dereddened stellar continuum normalized stacks. This assumes a clumpy geometry with a uniform dust screen. The blue lines show when the C_f(metal) is derived in the same geometry as C_f(H I). When calculated with consistent geometries, C_f(Si II 1260Å) (blue dashed line) is offset from C_f(H I) by a similar amount as our observations in Fig. 4. The green dashed line shows our fitted relationship between C_f(Si II) and C_f(H I) from Fig. 4, after converting it to the clumpy geometry. The Reddy et al. (2016b) C_f(H I)-C_f(Si II) relationship is consistent, within the errors, with ours. The red dashed line shows a one-to-one relation.

Reddy et al. (2016b). We correct the observed residual flux from Reddy et al. (2016b) to consistently recompute the C_f(metal) values using Eq. (15) and their measured $E_{\rm B-V}$ (using their SMC extinction law) and C_f(H I) values (Fig. 6). The large offset between the Si II 1260Å and H I is halved (compare the dashed blue line to the dashed black line), although C_f(metal) is still systematically smaller by ∼ 0.2. This offset is similar to the offset that we measured in Fig. 4 (green dashed line in Fig. 6). The discrepancy between C_f(metal) and C_f(H I) emphasizes that C_f can only be read off from the residual flux of absorption lines in the uniform dust screen geometry because this continuum normalization accounts for the dust attenuation. In other geometries, the dust attenuation must be accounted for when measuring C_f.

The small remaining offset (∼ 0.2) between the C_f(H I) and C_f(metal) is comparable to the C_f(Si II)-C_f(H I) relation found in Eq. (12). To demonstrate this, we convert our C_f(H I) and C_f(Si II) values, measured in a uniform geometry, to a clumpy geometry, and recalculate how C_f(Si II) scales with C_f(H I) (green line in Fig. 6). This recalculated relation is within the errors of the Reddy et al. (2016b) Si II 1260Å relation (blue dashed line). The consistent scaling between C_f(H I) and C_f(Si II) from our study and Reddy et al. (2016b) demonstrates first that C_f(Si II) ≠ C_f(H I), and second that empirical relations can convert C_f(Si II) into C_f(H I) for a variety of galaxy types. Both findings are crucial for indirect measurements of the Lyman continuum escape fraction, as discussed in Paper II.

There is still a C_f spread among the different transitions: strong Si II 1260Å and C II lines have similar C_f, but weaker Si II 1527 and Al III lines have smaller C_f. Puzzlingly, the Si II 1260Å and Si II 1527Å lines arise from the same transition, and should have the same C_f. Different C_f(Si II 1260Å) and C_f(Si II 1527Å) would be expected if Si II 1527Å was not saturated, but Reddy et al. (2016b) show that the Si II 1527Å/Si II 1260Å equivalent width ratio is unity. Consequently, both transitions are saturated. Alternatively, since Si II 1260Å is a stronger line than Si II 1527Å (f-value of 1.22 versus 0.133), there could be more optically thick Si II 1260Å regions along the line of sight than there are optically thick Si II 1527Å regions. Intersecting more regions along the line of sight covers more of the background continuum, such that C_f(Si II 1260Å) is larger than C_f(Si II 1527Å). This differential covering fraction is similar to the proposed physical mechanism creating the difference between C_f(Si II) and C_f(H I) in Sect. 5.3. In the Si II 1260Å and Si II 1527Å case, the number of absorbers is different due to differences in line strength, not metallicity.

6. Conclusions

We have analyzed the H I, O I, and Si II low ionization interstellar absorption lines from a sample of 18 star-forming galaxies with restframe ultraviolet spectroscopy of the Lyman series. The majority of the sources have COS spectra taken with the HST and are at low redshift ($z < 0.4$). Our sample includes nine Lyman continuum leaking galaxies. We fit the stellar continuum, dust attenuation, H I Lyman series absorption lines, as well as several low ionization absorption lines (Si II $\lambda\lambda$1190, 1193 and λ1260 Å, O I λ 1039 Å in particular). These fits determine the UV attenuation, as well as column densities and covering fractions of neutral hydrogen and metals (Sect. 3). Additionally, we applied our fitting method to synthetic Lyman series ISM absorption lines to investigate the systematic errors of the covering fractions and column densities. Then, we studied the observed Lyman series lines to constrain the H I properties. The direct H I properties were compared to indirect estimates of the neutral gas properties using O I and Si II absorption lines.

Our main results are summarized as follows:

- The H I covering fraction is accurately recovered from the Lyman series. Synthetic spectra recover the covering fractions with low systematics for a wide range of signal-to-noise ratios and resolutions (SNR $\geq$ 2 and $R > 3000$, or SNR $\geq$ 5 and $R \geq 600$; Fig. 2). Future observatories, like *JWST*, will accurately measure covering fractions of high-redshift galaxies.
- The observed H I lines are found to be saturated in all galaxies. Assuming the same O/H abundance in the neutral and ionized gas, we derive H I column densities of $\log(N_{\mathrm{H\,I}}) \sim 18.6 - 20$ cm^{-2} from the O I absorption lines (Sect. 3.3.2).
- The H I column densities derived for the known LyC leakers are too high to allow ionizing photons to escape. Rather, we find that the LyC emitting galaxies have H I covering fractions below unity. Ionizing photons escape through optically thin holes in a clumpy interstellar medium. The median H I covering fraction of confirmed LyC emitting galaxies is 0.62, as compared to 0.95 for galaxies that do not have LyC detections.
- The Si II covering fraction is systematically lower than the H I covering fraction (Fig. 4). However, the Si II covering fraction is found to scale linearly with the H I covering fraction (Eq. (12)). We show that this relation is compatible with the relationship from Reddy et al. (2016b) of stacked $z \sim 3$ spectra. Thus, with an empirical correction, the Si II absorption lines can be used to determine the H I covering fraction. This is especially powerful at high redshift when the Lyman series cannot be observed.
- The assumed dust geometry (here a uniform screen) impacts the measured covering fractions and dust attenuations (Fig. 5), but it does not impact the inferred escape fractions of ionizing photons (Sect. 5.3). Crucially, the geometric covering fraction cannot be read off from the residual line flux in all geometries. A consistent fitting of the dust attenuation, continuum, and absorption lines is required to properly determine the covering fraction.

By relating the H I covering fraction to metal absorption lines, we have provided the framework to measure the neutral gas covering fractions in star-forming galaxies. This will be of particular interest at high redshift, where the Lyman series and Lyman continuum are not observable and indirect measurements are required to measure the escape fraction of ionizing photons. In a companion paper (Chisholm et al. submitted), we use the absorption lines to accurately predict the escape fraction of the known LyC leakers. These methods also yield consistent results for a sample of $z \sim 2 - 2.4$ galaxies. Our analysis emphasizes that UV spectra of sufficient quality (SNR > 5) from *JWST* and extremely large telescopes may constrain how the Universe became reionized.

Acknowledgements.

References

Alexandroff, R. M., Heckman, T. M., Borthakur, S., Overzier, R., & Leitherer, C. 2015, ApJ, 810, 104
Bergvall, N., Zackrisson, E., Andersson, B.-G., et al. 2006, A&A, 448, 513
Bian, F., Fan, X., McGreer, I., Cai, Z., & Jiang, L. 2017, The Astrophysical Journal, 837, L12
Borthakur, S., Heckman, T. M., Leitherer, C., & Overzier, R. A. 2014, Science, 346, 216
Chisholm, J., Orlitová, I., Schaerer, D., et al. 2017a, A&A, 605, A67
Chisholm, J., Tremonti, C. A., Leitherer, C., & Chen, Y. 2017b, MNRAS, 469, 4831
Chisholm, J., Tremonti, C. A., Leitherer, C., Chen, Y., & Wofford, A. 2016, MNRAS, 457, 3133
Chisholm, J., Tremonti, C. A., Leitherer, C., et al. 2015, ApJ, 811, 149
de Barros, S., Vanzella, E., Amorín, R., et al. 2016, A&A, 585, A51
Dressler, A., Henry, A., Martin, C. L., et al. 2015, ApJ, 806, 19
Fontanot, F., Cristiani, S., Pfrommer, C., Cupani, G., & Vanzella, E. 2014, MNRAS, 438, 2097
Fontanot, F., Cristiani, S., & Vanzella, E. 2012, MNRAS, 425, 1413
Green, J. C., Froning, C. S., Osterman, S., et al. 2012, ApJ, 744, 60
Hamann, F., Barlow, T. A., & Junkkarinen, V. 1997, ApJ, 478, 87
Heckman, T. M., Borthakur, S., Overzier, R., et al. 2011, ApJ, 730, 5
Heckman, T. M., Lehnert, M. D., Strickland, D. K., & Armus, L. 2000, ApJS, 129, 493
Henry, A., Scarlata, C., Martin, C. L., & Erb, D. 2015, ApJ, 809, 19
Izotov, Y. I., Guseva, N. G., & Thuan, T. X. 2011, ApJ, 728, 161
Izotov, Y. I., Orlitová, I., Schaerer, D., et al. 2016a, Nature, 529, 178
Izotov, Y. I., Schaerer, D., Thuan, T. X., et al. 2016b, MNRAS, 461, 3683
Izotov, Y. I., Schaerer, D., Worseck, G., et al. 2018, MNRAS, 474, 4514
Jaskot, A. E. & Oey, M. S. 2013, ApJ, 766, 91
Jones, T. A., Ellis, R. S., Schenker, M. A., & Stark, D. P. 2013, ApJ, 779, 52
Koester, B. P., Gladders, M. D., Hennawi, J. F., et al. 2010, ApJ, 723, L73
Lebouteiller, V., Heap, S., Hubeny, I., & Kunth, D. 2013, A&A, 553, A16
Leitet, E., Bergvall, N., Hayes, M., Linné, S., & Zackrisson, E. 2013, A&A, 553, A106
Leitherer, C., Hernandez, S., Lee, J. C., & Oey, M. S. 2016, ApJ, 823, 64

Leitherer, C., Ortiz Otálvaro, P. A., Bresolin, F., et al. 2010, ApJS, 189, 309
Leitherer, C., Schaerer, D., Goldader, J. D., et al. 1999, ApJS, 123, 3
Markwardt, C. B. 2009, in Astronomical Society of the Pacific Conference Series, Vol. 411, Astronomical Data Analysis Software and Systems XVIII, ed. D. A. Bohlender, D. Durand, & P. Dowler, 251
Marshall, J. L., Burles, S., Thompson, I. B., et al. 2008, in Proc. SPIE, Vol. 7014, Ground-based and Airborne Instrumentation for Astronomy II, 701454
Meynet, G., Maeder, A., Schaller, G., Schaerer, D., & Charbonnel, C. 1994, A&AS, 103, 97
Nakajima, K. & Ouchi, M. 2014, MNRAS, 442, 900
O'Meara, J. M., Prochaska, J. X., Burles, S., et al. 2007, ApJ, 656, 666
Ouchi, M., Mobasher, B., Shimasaku, K., et al. 2009, ApJ, 706, 1136
Pettini, M. & Pagel, B. E. J. 2004, MNRAS, 348, L59
Pettini, M., Rix, S. A., Steidel, C. C., et al. 2002, ApJ, 569, 742
Prochaska, J. X., Kasen, D., & Rubin, K. 2011, ApJ, 734, 24
Puschnig, J., Hayes, M., Östlin, G., et al. 2017, ArXiv e-prints
Quider, A. M., Shapley, A. E., Pettini, M., Steidel, C. C., & Stark, D. P. 2010, MNRAS, 402, 1467
Reddy, N. A., Steidel, C. C., Pettini, M., & Bogosavljević, M. 2016a, ApJ, 828, 107
Reddy, N. A., Steidel, C. C., Pettini, M., Bogosavljević, M., & Shapley, A. E. 2016b, ApJ, 828, 108
Rigby, J. R., Bayliss, M. B., Sharon, K., et al. 2017, ArXiv e-prints
Robertson, B. E., Furlanetto, S. R., Schneider, E., et al. 2013, ApJ, 768, 71
Scarlata, C. & Panagia, N. 2015, ApJ, 801, 43
Shapley, A. E., Steidel, C. C., Pettini, M., & Adelberger, K. L. 2003, ApJ, 588, 65
Shapley, A. E., Steidel, C. C., Strom, A. L., et al. 2016, ApJ, 826, L24
Stark, D. P., Swinbank, A. M., Ellis, R. S., et al. 2008, Nature, 455, 775
Trainor, R. F., Steidel, C. C., Strom, A. L., & Rudie, G. C. 2015, ApJ, 809, 89
Vanzella, E., de Barros, S., Castellano, M., et al. 2015, A&A, 576, A116
Vanzella, E., Siana, B., Cristiani, S., & Nonino, M. 2010, MNRAS, 404, 1672
Vasei, K., Siana, B., Shapley, A. E., et al. 2016, ApJ, 831, 38
Wakker, B. P., Hernandez, A. K., French, D. M., et al. 2015, ApJ, 814, 40
Weiner, B. J., Coil, A. L., Prochaska, J. X., et al. 2009, ApJ, 692, 187
Worseck, G., Prochaska, J. X., Hennawi, J. F., & McQuinn, M. 2016, The Astrophysical Journal, 825, 144
Wuyts, E., Rigby, J. R., Gladders, M. D., et al. 2012, ApJ, 745, 86
Zackrisson, E., Inoue, A. K., & Jensen, H. 2013, ApJ, 777, 39

Appendix A: Simulated spectra

Here we detail the procedure we used to create the simulated spectra in Sect. 3.2. We simulate synthetic spectra to mimic the observed Lyman series. Therefore, we create synthetic spectra using a single STARBURST99 stellar continuum model of R(S99) ≈ 2500 (Leitherer et al. 1999, 2010) with a stellar age of 3 Myr, a solar metallicity, and without dust attenuation. This stellar age has a strong O IV P-Cygni profile (broad blueshifted absorption and redshifted emission) which blends with the Lyβ absorption line (see Fig. 1). The synthetic spectra include the wavelength regime between 910-1050 Å with a pixel separation of 0.02 Å (6 km/s at 1035 Å). These wavelengths include the Lyman series, from Lyβ to the Lyman break (≈ 911.8 Å). We imprint on top of the stellar continuum interstellar medium absorption lines from H I, O I, O VI and C II (see Table 2) with Voigt profiles. The Doppler broadening parameters (b), column densities (N), and covering fractions (C_f) are created with independent values for each element (see Table A.1). We fixed the velocities of each line at 0 km s^{-1}. The Einstein-A coefficients and the oscillator strengths of each transition were taken from the NIST Atomic Database[2].

We define two categories of synthetic spectra corresponding to an ionizing escape fraction of 10%. This value is drawn from the largest known publicly available low-redshift LyC escape fractions (Izotov et al. 2016b,a). For the first category, we assume holes in an otherwise optically thick ISM allow ionizing photons to escape (we call this the "picket-fence" model). For the picket-fence model, the H I parameters are $N_{\rm H\,I}$ = 10^{20} cm^{-2} ($e^{-\tau_H} \rightarrow 0$) and C_f(H I) = 0.90. For the second scenario, we assume a low H I column density allows ionizing photons to escape (the "density-bounded" regime). We fix $N_{\rm H\,I}$ = $10^{17.57}$ cm^{-2} ($e^{-\tau_H} \approx 0.10$) and C_f(H I) = 1.0. Since we fix $E_{\rm B-V} = 0$ in all our simulations, the geometry does not change our simulations (Sect. 3.1.1).

To reproduce the observations, we create synthetic spectra at different instrumental resolutions and SNRs. We simulate seven resolutions, $R = \lambda/\delta\lambda$ (where $\delta\lambda$ is the FWHM of the spectra), of 15000, 6000, 3000, 1500, 1000, 750 and 600. For the synthetic spectra having R > R(S99), this is done by combining the theoretical stellar continuum with R(S99) = 2500 with absorption lines of spectral resolution R. The range of resolutions chosen correspond to spectrographs like HST/COS, LRIS/HRIS on Keck, MagE on Magellan, Muse, or the upcoming NIRSpec on *JWST*. For each R, we generate seven SNRs (2, 5, 10, 20, 30, 40 and 50) in a sufficiently large range to study typically delivered noise levels. At each SNR, we generate 50 different sets of random Gaussian noise and add the noise to the same synthetic spectra. This produces a sample of 50 synthetic spectra where only the random noise varies between the spectra. Figures B.1 and B.2 show simulated spectra with resolutions of 15000 and 600 for three SNRs (noise-free, 10 and 2). These results are discussed in Sect. 4.

Table A.1: Absorption lines, wavelengths, and parameters used to create the synthetic absorption lines.

(1) Ion	(2) v [km s^{-1}]	(3) b [km s^{-1}]	(4) log(N) [cm^{-2}]	(5) C_f
H I	0	50	17.57[a]	1[a]
			20.00[b]	0.9[b]
O I	0	75	16	1
O VI	0	100	15	1
C II	0	80	15.5	1

Notes. (1) Ion; (2) velocity shift of the lines; (3) Doppler parameter; (4) logarithm of the column density; (5) covering fraction.
[(a)] Denotes values used for the simulated "density-bounded" spectra
[(b)] Denotes values used for the simulated "Picket-Fence" spectra

Appendix B: Fits

Here we provide fits of the synthetic and observed spectra detailed in this article. Fig. B.1 and Fig. B.2 show the synthetic spectra for three SNRs (∞, 10 and 2) with R = 15000 and R = 600, respectively. In Figs B.3–B.20 we present the fits for the 18 galaxies in our sample. The top panels of each figure show the full wavelength coverage, with the observed flux in black and the total fit in red. Gray flux indicates that these regions were masked during the fit. The green lines show the error on the observed flux. The lower panels zoom in on individual Lyman series lines. The upper portions of each plot indicate the fitted Lyman series lines (blue), ISM metal absorption lines (solid black line), and Milky Way absorption lines (black dashed lines). The line labels are gray if they are not fit. Geocoronal regions are denoted by shaded gray regions.

[2] https://physics.nist.gov/PhysRefData/ASD/lines_form.html

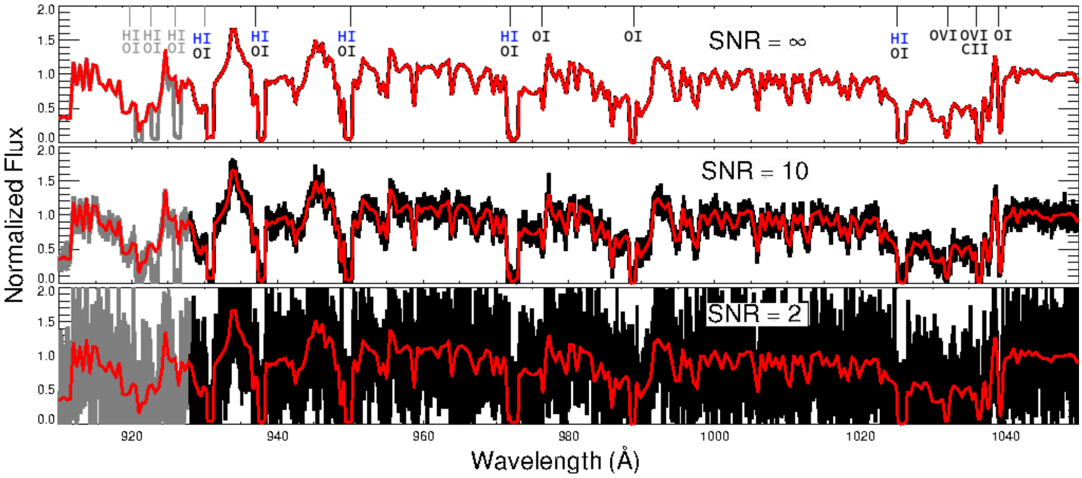

Fig. B.1: Fits (in red) of the synthetic spectra at high resolution (R = 15000) for the Picket-Fence model (C_f(H I) = 0.9, $N_{\mathrm{H\,I}}$= 10^{20}cm^{-2}) with SNR equal to ∞ (top), 10 (middle) and 2 (bottom). We exclude the gray portion when fitting. In the upper portion of the top panel, we indicate and label the fitted absorption lines.

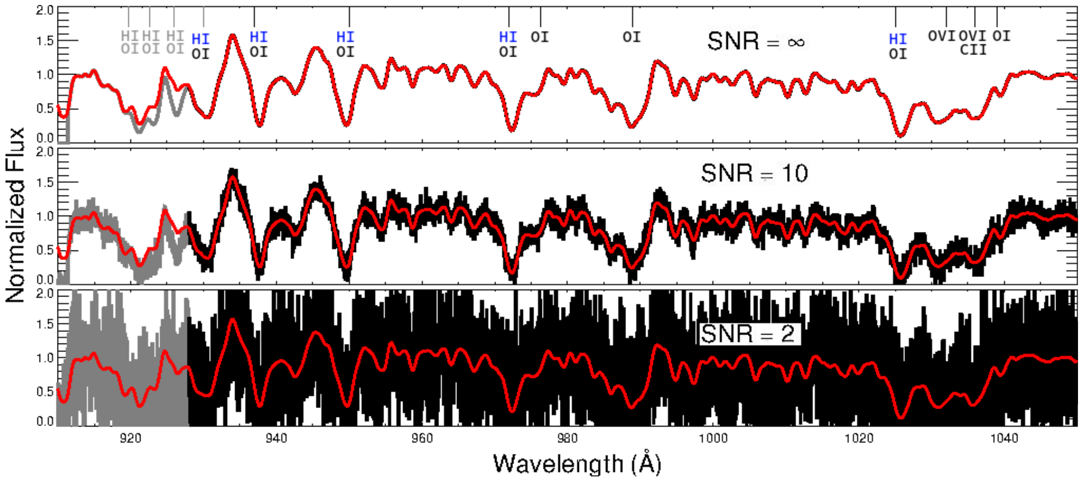

Fig. B.2: Same as Fig. B.1 but for a spectral resolution of R = 600.

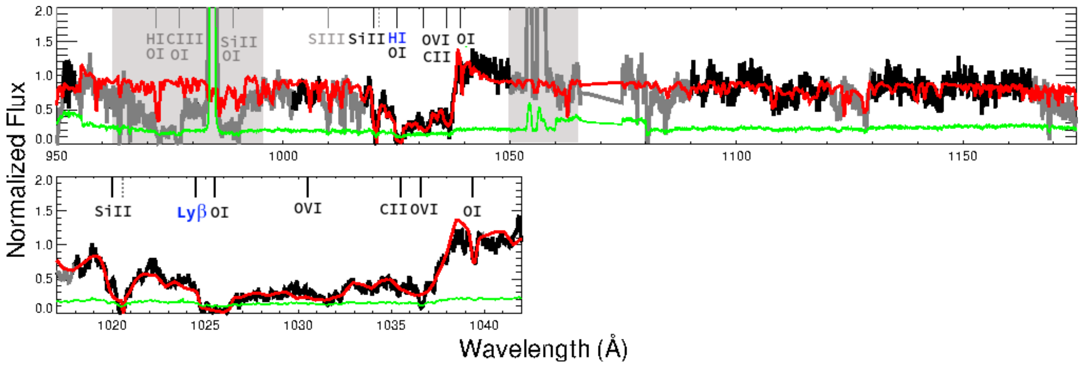

Fig. B.3: Fit (in red solid line) of the COS G130M spectrum of the galaxy J0921+4509. Black is the observed flux included in the routine either to fit the stellar continuum or the ISM absorption lines. Gray portions are masked out for both steps; grey shaded regions indicate those masked because of geocoronal emission. The flux error array appears in green. We display the ISM and Milky Way absorption lines as solid and dotted lines in the upper portion of each panel. Black or blue labels indicate that the lines are fit, whereas gray labels indicate that they are not fit. When present, red labels indicate lines which are not detected. Reference for this observation: Borthakur et al. (2014)

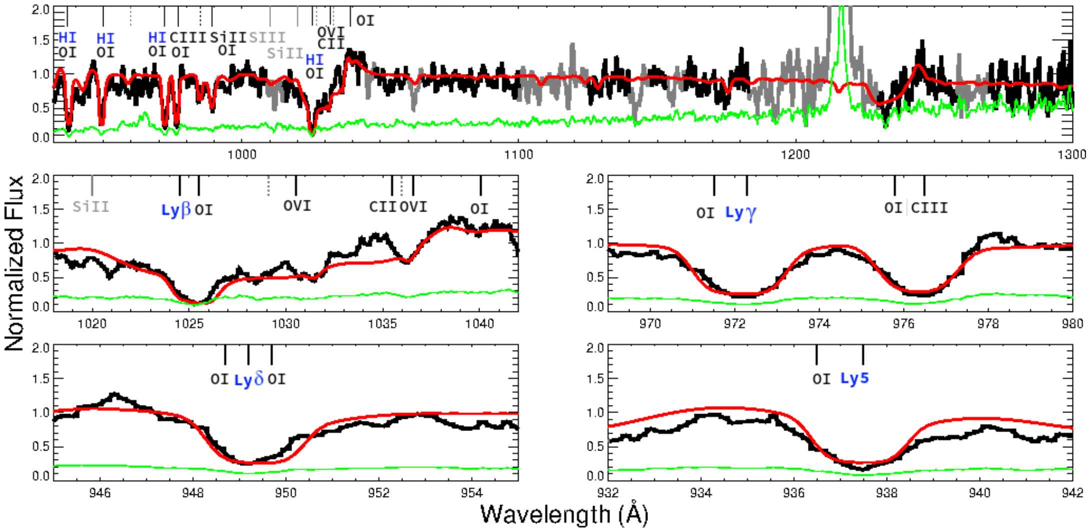

Fig. B.4: Same as Fig. B.3 but for the COS G140L spectrum of J1503+3644. Reference for this observation: Izotov et al. (2016b)

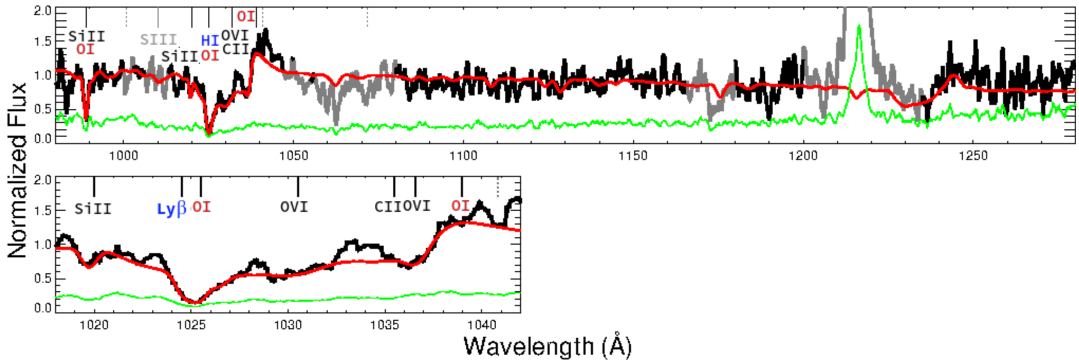

Fig. B.5: Same as Fig. B.3 but for the COS G140L spectrum of J0925+1409. Reference for this observation: Izotov et al. (2016a)

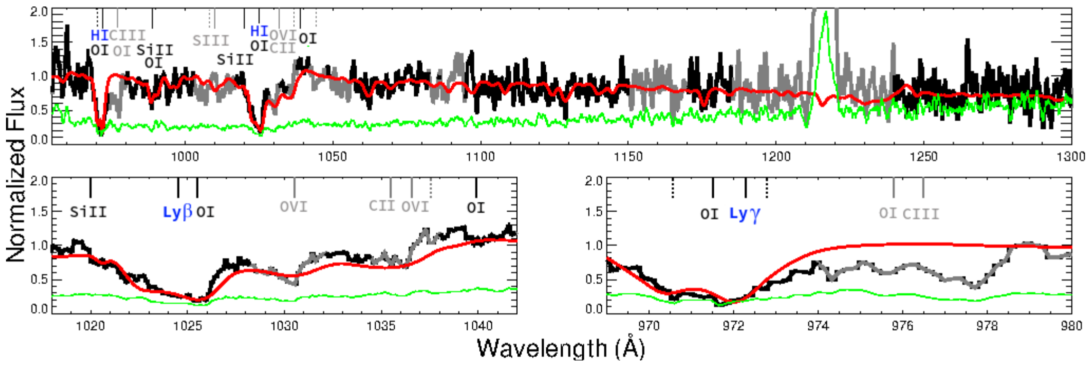

Fig. B.6: Same as Fig. B.3 but for the COS G140L spectrum of J1152+3400. Reference for this observation: Izotov et al. (2016b)

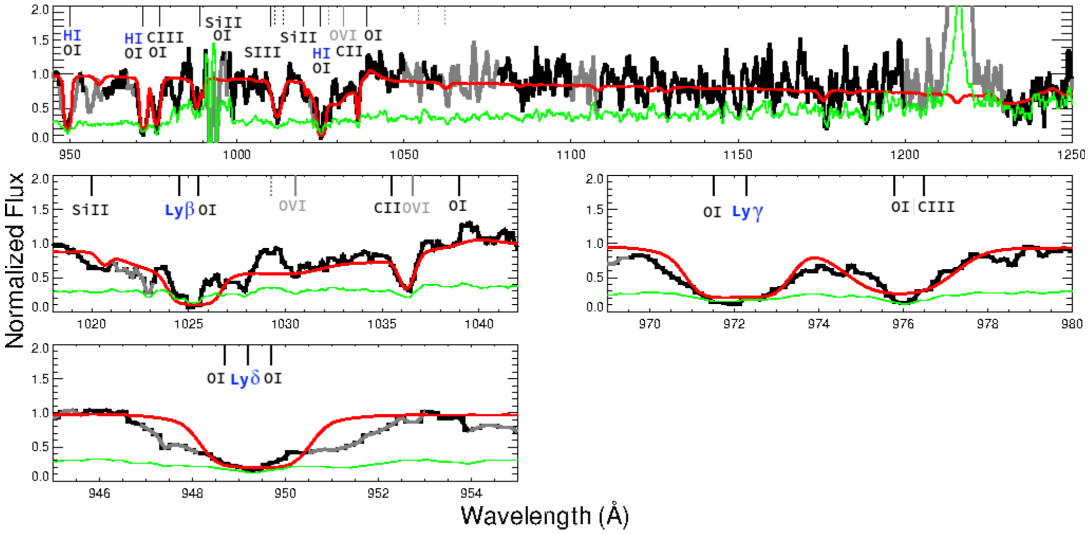

Fig. B.7: Same as Fig. B.3 but for the COS G140L spectrum of J1333+6246. Reference for this observation: Izotov et al. (2016b)

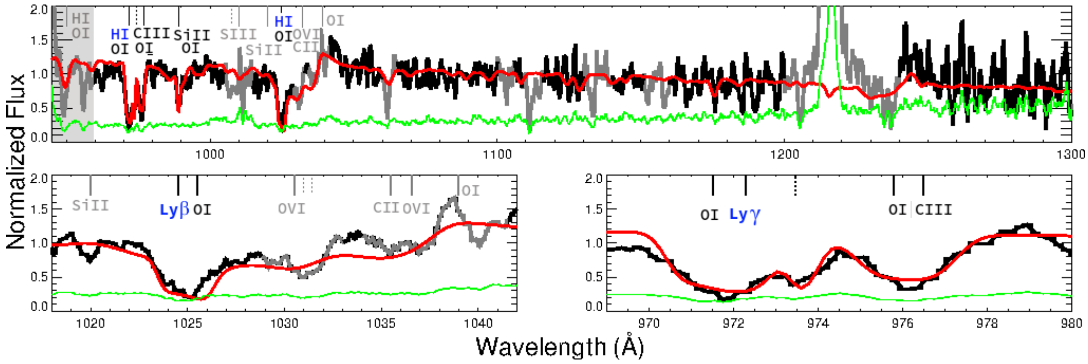

Fig. B.8: Same as Fig. B.3 but for the COS G140L spectrum of J1442-0209. Reference for this observation: Izotov et al. (2016b)

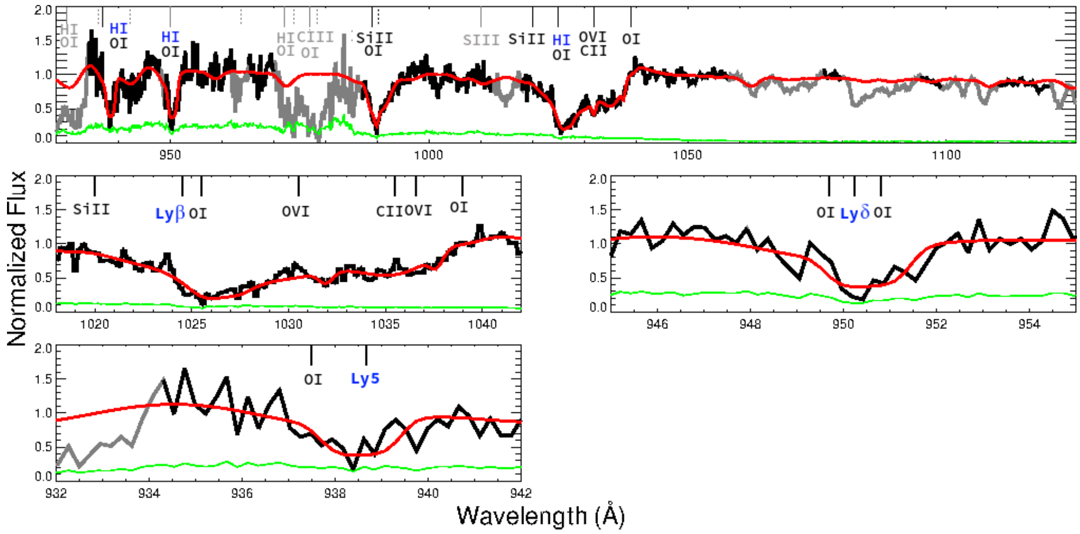

Fig. B.9: Same as Fig. B.3 but for the COS G140L spectrum of Tol1247-232. Reference for this observation: Leitherer et al. (2016)

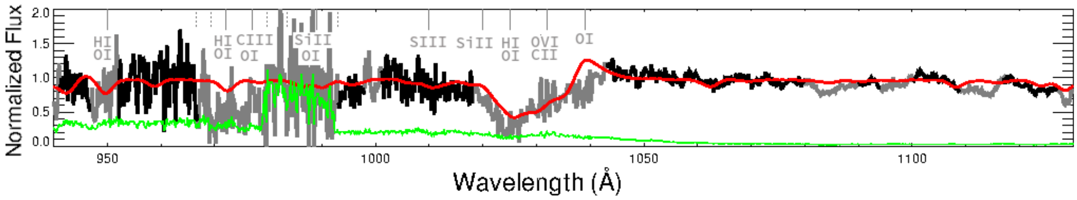

Fig. B.10: Best fit for the stellar continuum for the COS G140L spectrum of Tol0440-381. This galaxy has a low-redshift, and the Milky Way absorption lines contaminate the Lyman series. Consequently, we do not fit for the ISM absorption lines. Reference for this observation: Leitherer et al. (2016)

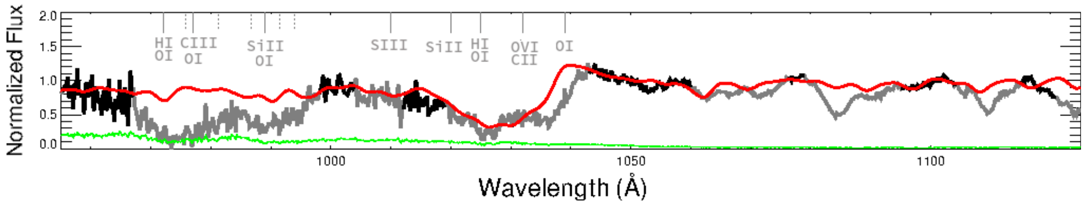

Fig. B.11: Same as Fig. B.10 but for the COS G140L spectrum of Mrk54. Reference for this observation: Leitherer et al. (2016)

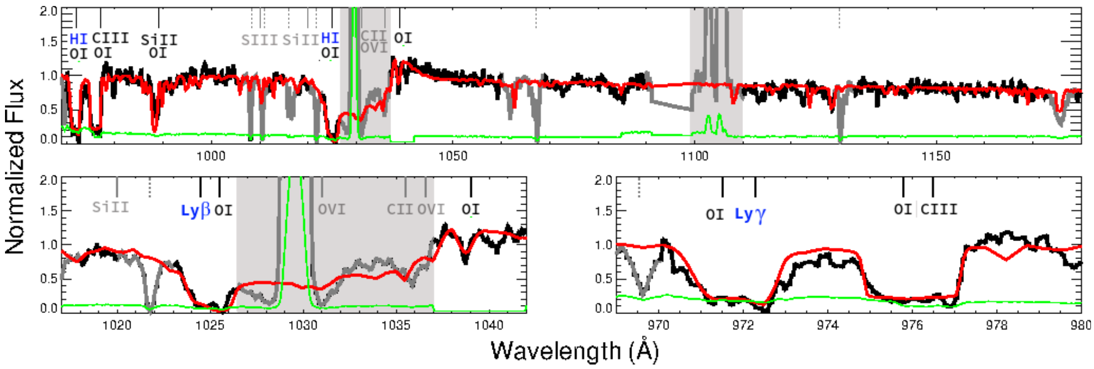

Fig. B.12: Same as Fig. B.3 but for the COS G130M spectrum of J0926+4427. Reference for this observation: Heckman et al. (2011)

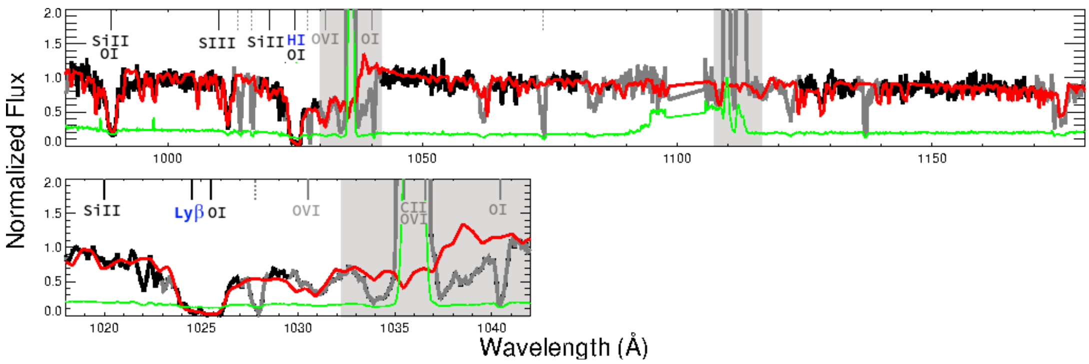

Fig. B.13: Same as Fig. B.3 but for the COS G130M spectrum of J1429+0643. Reference for this observation: Heckman et al. (2011)

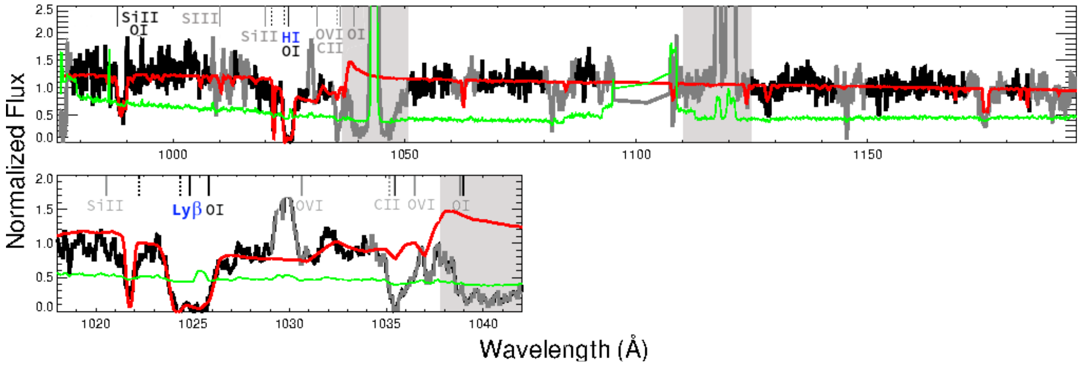

Fig. B.14: Same as Fig. B.3 but for the COS G130M spectrum of GP0303-0759. Reference for this observation: Henry et al. (2015)

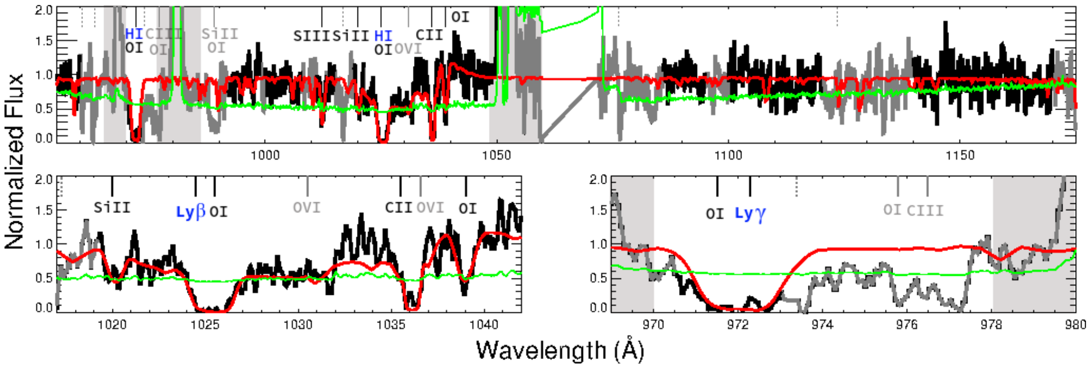

Fig. B.15: Same as Fig. B.3 but for the COS G130M spectrum of GP1244+0216. Reference for this observation: Henry et al. (2015)

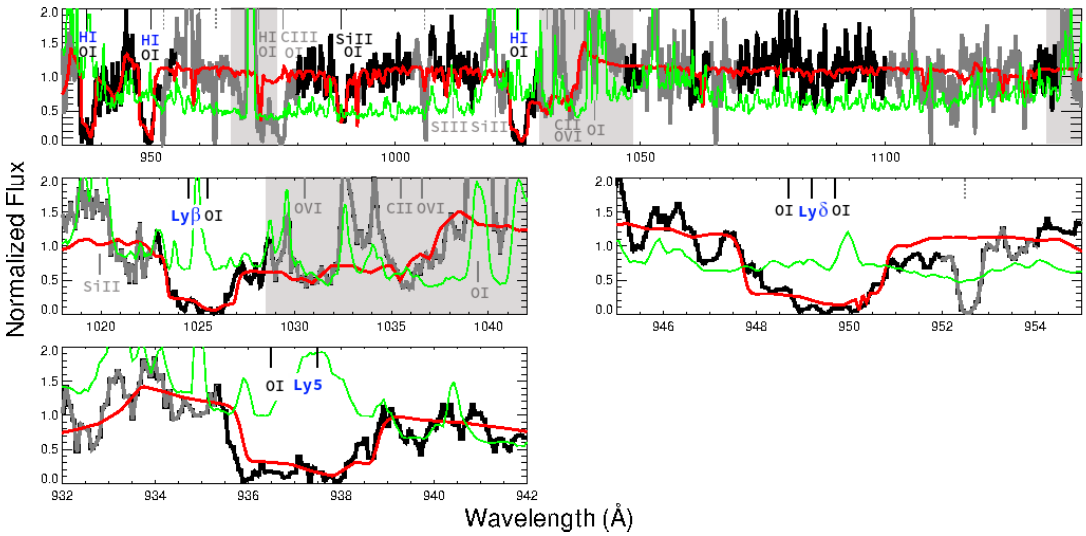

Fig. B.16: Same as Fig. B.3 but for the COS G130M spectrum of GP1054+5238. Reference for this observation: Henry et al. (2015)

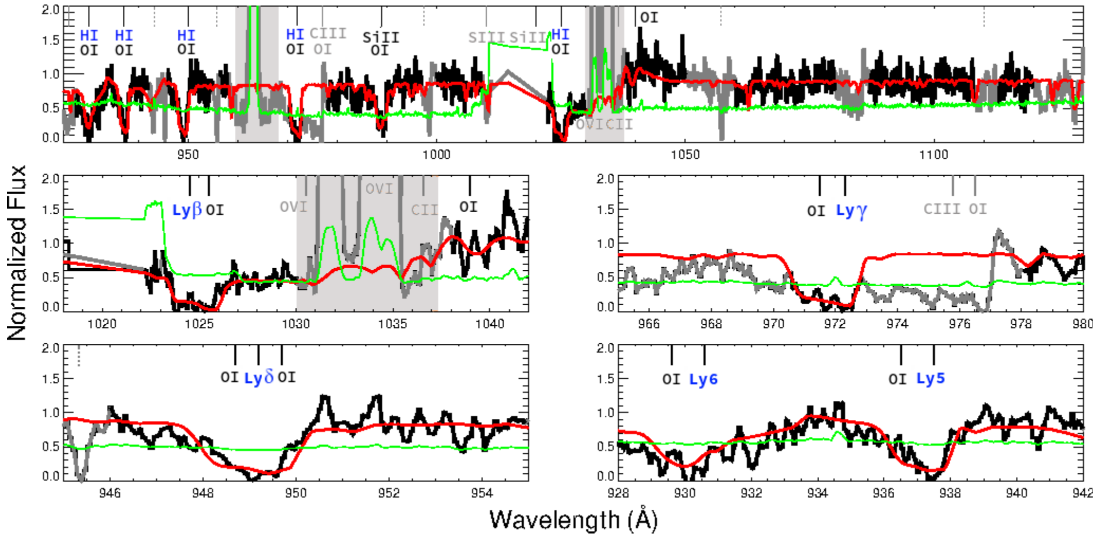

Fig. B.17: Same as Fig. B.3 but for the COS G130M spectrum of GP0911+1831. Reference for this observation: Henry et al. (2015)

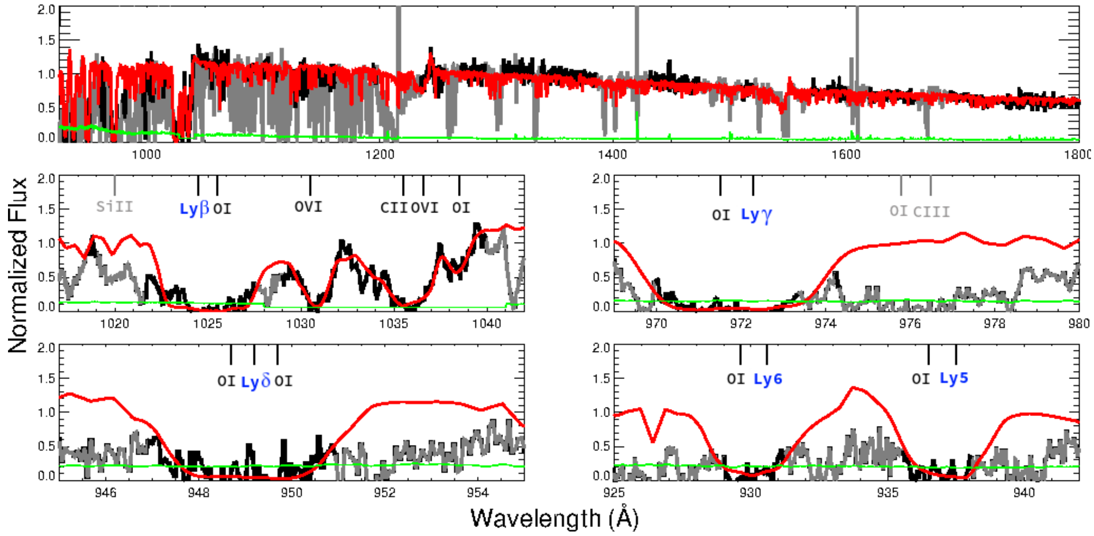

Fig. B.18: Same as Fig. B.3 but for the MagE spectrum of SGAS J122651.3+215220. Reference for this observation: Rigby et al. (2017)

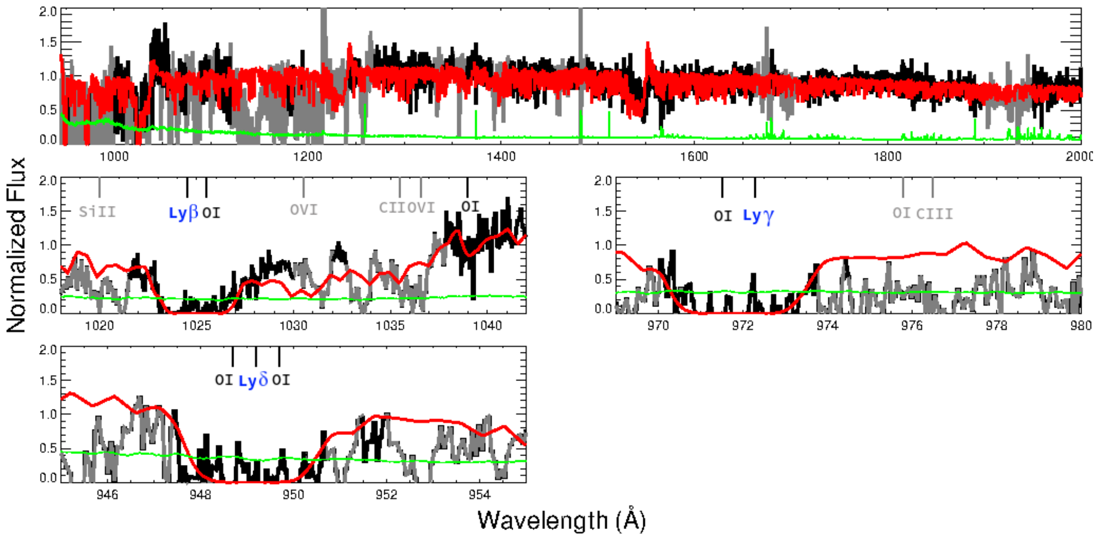

Fig. B.19: Same as Fig. B.3 but for the MagE spectrum of SGAS J152745.1+065219. Reference for this observation: Rigby et al. (2017)

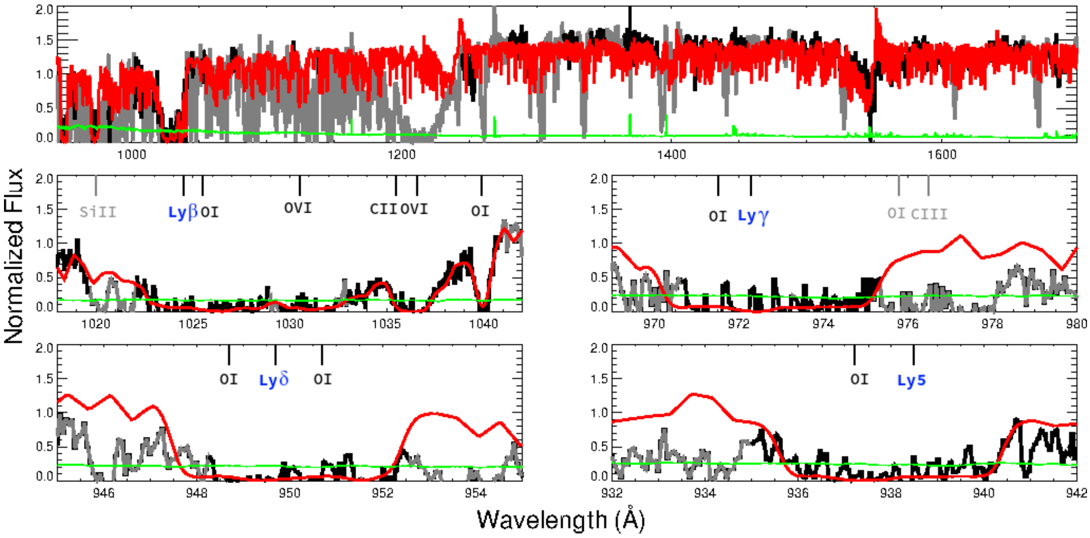

Fig. B.20: Same as Fig. B.3 but for the MagE spectrum of the Cosmic Eye. Reference for this observation: Rigby et al. (2017)

Appendix C: Tables

Appendix C.1: Measured systematic errors from single synthetic spectra

Here we tabulate the systematic errors measured from the synthetic spectra (Sect. 4). Table C.1 gives the $N_{\mathrm{H\,I}}$ systematic errors, and Table C.2 gives the C_f systematic errors. The percent errors plotted in Fig. 2 are computed by subtracting 0.9 from the errors and dividing the difference by 0.9 (the C_f used to generate the spectra).

Appendix C.2: Lyman series residual flux

Table C.3 lists the residual flux of the individual Lyman series lines (Sect. 3.3.1).

Table C.1: Systematic errors on the logarithm of the H I column density estimated from synthetic spectra.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
R	Signal-to-noise ratio						
	50	40	30	20	10	5	2
15000	0.04	0.04	0.05	0.07	0.17	0.33	1.13
6000	0.05	0.07	0.10	0.14	0.31	0.78	1.27
3000	0.09	0.09	0.12	0.18	0.57	1.05	1.13
1500	0.10	0.17	0.20	0.28	0.65	1.04	1.28
1000	0.12	0.14	0.17	0.30	0.87	1.21	1.18
750	0.14	0.17	0.24	0.33	0.86	1.20	1.53
600	0.13	0.19	0.25	0.30	0.81	1.21	1.54

Notes. Columns 2-8 give the systematic $\log(N_{\mathrm{H\,I}})$ error at different spectral resolutions (R; see column 1) and signal-to-noise ratios (see columns 2-7 for SNR 50-2) for simulated spectra with $N_{\mathrm{H\,I}} = 10^{17.57}\mathrm{cm}^{-2}$. See Sect. 4.1.

Table C.2: Systematic errors on the covering fraction estimated from synthetic spectra.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
R	Signal-to-noise ratio						
	50	40	30	20	10	5	2
15000	<0.01	<0.01	<0.01	0.01	0.01	0.02	0.06
6000	<0.01	<0.01	<0.01	0.01	0.01	0.03	0.07
3000	<0.01	<0.01	<0.01	0.01	0.01	0.03	0.07
1500	<0.01	0.01	0.01	0.01	0.03	0.05	0.10
1000	0.01	0.011	0.02	0.02	0.04	0.07	0.12
750	0.02	0.021	0.03	0.04	0.06	0.09	0.14
600	0.03	0.032	0.04	0.05	0.09	0.10	0.17

Notes. Columns 2-8 give the systematic C_f errors at different spectral resolutions (R; column 1) and signal-to-noise ratios (see columns 2-7 for SNR 50-2) for simulated spectra with $C_f = 0.9$. See Sect. 4.2.

Table C.3: Measurement of H I covering fraction derived from the residual flux of the individual Lyman series. See Sect. 3.3.1

Galaxy name	Lyβ	Lyγ	Lyδ	Ly5	Weighted mean
J0921+4509	0.769 ± 0.116	-	-	-	0.769 ± 0.116
J1503+3644	0.847 ± 0.157	0.723 ± 0.129	0.741 ± 0.112	0.744 ± 0.113	0.754 ± 0.062
J0925+1409	0.635 ± 0.094	-	-	-	0.635 ± 0.094
J1152+3400	0.619 ± 0.088	-	-	-	0.619 ± 0.088
J1333+6246	0.773 ± 0.152	0.870 ± 0.103	0.804 ± 0.106	-	0.826 ± 0.066
J1442-0209	0.589 ± 0.049	0.471 ± 0.068	-	-	0.549 ± 0.040
Tol1247-232	0.543 ± 0.135	-	0.775 ± 0.134	0.791 ± 0.175	0.690 ± 0.084
Tol0440-381	0.507 ± 0.139	-	0.615 ± 0.167	0.602 ± 0.137	0.570 ± 0.084
Mrk54	0.397 ± 0.136	-	0.652 ± 0.120	0.406 ± 0.148	0.504 ± 0.077
J0926+4427	0.817 ± 0.057	0.807 ± 0.087	-	-	0.814 ± 0.048
J1429+0643	0.955 ± 0.061	-	-	-	0.955 ± 0.061
GP0303-0759	-	-	-	-	-
GP1244+0216	0.985 ± 0.211	0.894 ± 0.204	1.000 ± 0.292	-	0.950 ± 0.131
GP1054+5238	0.936 ± 0.318	-	0.891 ± 0.203	0.798 ± 0.420	0.889 ± 0.158
GP0911+1831	0.718 ± 0.361	0.781 ± 0.189	0.825 ± 0.198	0.635 ± 0.282	0.765 ± 0.116
SGAS J1226	1.000 ± 0.010	1.000 ± 0.030	0.981 ± 0.057	0.940 ± 0.060	0.998 ± 0.009
SGAS J1527	0.858 ± 0.141	1.000 ± 0.045	1.000 ± 0.087	1.000 ± 0.201	0.990 ± 0.038
Cosmic Eye	1.000± 0.043	1.000 ± 0.063	1.000 ± 0.033	0.899 ± 0.166	0.998 ± 0.024

Notes. Dashes indicate that these transitions were not observed due to Milky Way absorption, geocoronal emission, or low SNR.