The Lyman continuum escape fraction and the Mean Free Path of hydrogen ionizing photons for bright $z \sim 4$ QSOs from SDSS DR14

M. Romano[1], A. Grazian[2,3], E. Giallongo[3], S. Cristiani[4], F. Fontanot[4], K. Boutsia[5], F. Fiore[4], and N. Menci[3]

[1] Dipartimento di Fisica e Astronomia, Università di Padova, Vicolo dell'Osservatorio 3, I-35122, Padova, Italy
[2] INAF–Osservatorio Astronomico di Padova, Vicolo dell'Osservatorio 5, I-35122, Padova, Italy
[3] INAF–Osservatorio Astronomico di Roma, Via Frascati 33, I-00078, Monte Porzio Catone, Italy
[4] INAF–Osservatorio Astronomico di Trieste, Via G.B. Tiepolo 11, I-34143, Trieste, Italy
[5] Las Campanas Observatory, Carnegie Observatories, Colina El Pino Casilla 601, La Serena, Chile

ABSTRACT

Context. One of the major challenges in observational cosmology is related to the redshift evolution of the average hydrogen (HI) ionization in the Universe, as evidenced by the changing in the ionization level of the intergalactic medium (IGM) through cosmic time. In particular, starting from the first cosmic reionization, the rapid evolution of the IGM physical properties poses severe constraints for the identification of the sources responsible for keeping its high level of ionization up to lower redshifts.
Aims. In order to probe the ionization level of the IGM and the ionization capabilities of bright QSOs at $z = 4$, we have selected a sample of 2508 QSOs drawn from the SDSS survey (DR14) in the redshift interval $3.6 \leq z \leq 4.6$ and absolute magnitude range $-29.0 \lesssim M_{1450} \lesssim -26.0$. Particularly, we focus on the estimation of the escape fraction of HI ionizing photons and their mean free path (MFP), which are fundamental in the characterization of the surrounding intergalactic medium.
Methods. Starting from UV/optical rest-frame spectra of the whole QSO sample from the SDSS survey, we estimate the escape fraction and free path individually for each of them. We calculate the Lyman Continuum escape fraction as the flux ratio blue-ward (~ 900 Å rest-frame) and red-ward (~ 930 Å rest-frame) of the Lyman limit (LL). We then obtain the probability distribution function (PDF) of the individual free paths of the QSOs in the sample and study its evolution in luminosity and redshift, comparing our results with the ones found in literature so far.
Results. We find a lower limit to the mean Lyman Continuum escape fraction of 0.49, in agreement with the values obtained for both brighter and fainter sources at the same redshift. We show that the free paths of ionizing photons are characterized by a skewed distribution function peaked at low values, with an average of $\sim 49 - 59$ proper Mpc at $z \sim 4$, after excluding possible associated absorbers. This value is larger than the one obtained at the same redshift by many authors in the literature using different techniques. Moreover, the probability distribution function of free path gives a complementary information w.r.t. the mean free path derived through the stack technique. Finally, we also find that the redshift evolution of this parameter results to be possibly milder than previously thought.
Conclusions. Our new determination of the mean free path at $z \sim 4$ implies that previous estimates of the HI photo-ionization rate Γ_{HI} available in the literature should be corrected by a factor of 1.2-1.7. These results have important implications if extrapolated at the epoch of reionization.

Key words. quasars: general - Cosmology: reionization

1. Introduction

The epoch of reionization (EoR) marks a fundamental event for the Universe, characterized by a transition phase of the intergalactic medium (IGM) from cold and almost neutral to warm and fully ionized (Meiksin 2009; McQuinn 2016). During this dramatic event, located approximately at $z \sim 6 - 8$, the first sources of UV photons with energy above 13.6 eV were able to clear the fog of the widespread neutral hydrogen (HI) and put an end to the so called period of the Dark Ages (Dayal & Ferrara 2018). Only in the recent years, thanks mainly to the analysis of the CMB optical depth by Planck (Planck Collaboration 2018) and the high-z QSOs (Fan et al. 2006; Becker et al. 2015), it was clear that the EoR was a very rapid event; this phase transition lasted for a short time period $\Delta z \leq 2.8$, and it also was a patchy process. This is consistent with the progressive decrease

of the photo-ionization rate Γ_{HI} observed at $z \geq 5.5$ (Calverley et al. 2011; Davies et al. 2018; D'Aloisio et al. 2018).

A long-lasting debate is in progress on the sources responsible for the EoR. While the majority of the astrophysical community are favoring star-forming galaxies as the main drivers of HI reionization (Dayal & Ferrara 2018), alternative explanations based on Active Galactic Nuclei (AGNs) have been proposed (Giallongo et al. 2012, 2015; Madau & Haardt 2015; Grazian et al. 2018; Boutsia et al. 2018). It is widely accepted that QSOs and AGNs dominate the HI ionizing background at $z < 3$ (Haardt & Madau 2012; Fontanot et al. 2012; Cristiani et al. 2016; Faucher-Giguère 2019). At higher redshifts, however, the debate is still on-going whether the HI ionizing background is mainly driven by rest-frame star-forming galaxies or by AGNs. One of the main criticism against HI reionization driven by accreting super massive black holes (SMBHs) is the fast drop of the space density of bright QSOs (Fan et al. 2001; Cowie et al. 2009) at $z \geq 3$ and the lack of numerous $L \sim L^*$ AGNs at $z \geq 4$ (Parsa et

al. 2018; Akiyama et al. 2018; Kim et al. 2019; Matsuoka et al. 2018; Yang et al. 2018). In fact, contrasting results have been obtained in literature so far, leading to significant uncertainties in the estimation of the HI photo-ionization rate, in which the luminosity function directly enters (Boutsia et al. 2018; Giallongo et al. 2019; Kulkarni et al. 2019).

In order to quantify the contribution of active SMBHs to the cosmological photo-ionizing background, two other physical parameters are needed, in addition to the knowledge of the QSO luminosity function: 1-the escape fraction of HI ionizing photons, i.e. at the Lyman continuum (LyC) $\lambda \sim 900$ Å rest-frame, f_{esc}(LyC), and 2-the mean free path (MFP) of HI ionizing photons. The latter is the average physical distance HI-ionizing photons typically travel before being absorbed by a factor $1/e$, i.e. by an optical depth of $\tau_{HI} = 1$. This quantity directly enters into the calculation of the photo-ionization rate Γ_{HI}, and it can thus change in a significant way the amount of UV photons per unit time released by QSOs into the IGM.

At $z \sim 1$, Cowie et al. (2009) find f_{esc}(LyC) $\sim 100\%$ for bright QSOs, while Stevans et al. (2014) indicate similar results for a sample of fainter Seyfert 1 and Seyfert 2 galaxies. The LyC escape fraction of bright QSOs ($M_{1450} \lesssim -26$) and faint AGNs ($M_{1450} \sim -24$) at $z \sim 4$ has been studied by Cristiani et al. (2016) and Grazian et al. (2018), respectively. They find that the Lyman Continuum escape fraction of QSOs and AGNs is $\sim 75\%$ or greater, and it does not show any dependence on the absolute luminosities of the objects, at least up to $L \sim L^*$. In particular, Cristiani et al. (2016) find that the distribution of the escape fraction of bright QSOs from SDSS is bimodal, with 18-25% probability of negligible escape fraction and the remaining 75-82% with f_{esc}(LyC) $\sim 100\%$. However, as discussed in Grazian et al. (2018), this bi-modality could be due to their choice of the spectral window adopted to measure the ionizing photon leakage, i.e. $\lambda_{rest} = 865 - 885$ Å. Indeed, sources with high escape fraction and a proximate Lyman Limit System (LLS) or a Damped Lyman-α system (DLA), close to the emission redshift of a QSO, can spuriously mimic an f_{esc}(LyC) ~ 0. It is thus important to study in detail the relation between the LyC escape fraction and the presence of nearby absorbers. An effective parameter which is useful to quantify the presence of LLSs surrounding QSOs could be the mean free path.

The individual free path (IFP) of HI ionizing photons can be obtained by deriving the optical depth along a line of sight by counting the individual intervening absorbers (Faucher-Giguère et al. 2008; Songaila & Cowie 2010; Rudie et al. 2013; O'Meara et al. 2013; Inoue et al. 2014; Prochaska et al. 2015; Crighton et al. 2019). This approach yielded different results depending on the input HI column density distribution used in the models. The robust detection of absorbers, and in particular (partial) LLSs however, is a delicate issue and it has been limited to the line of sight of very luminous QSOs. They can reside in very massive halos, thus biasing our view of the MFP with respect to the one in the other, less biased, regions of the Universe. Alternatively, the mean free path can be obtained by stacking a large number of high-z QSOs and measuring the decrement of a factor $1/e$ with respect to the mean flux level at $\lambda_{rest} \sim 912$ Å, as carried out by Prochaska et al. (2009); Worseck et al. (2014). They obtain a fast evolution of the MFP with the cosmic epoch MFP(z) $\propto (1 + z)^{-5.4}$, over a large redshift range $2.3 < z < 5.5$. This rapid decrement implies, according to Worseck et al. (2014), an evolution in number density or on the physical size of the absorbers.

These two methods (individual absorber distribution vs stacking), however, are not providing fully consistent results. In particular, the redshift evolution of the MFP derived by Worseck et al. (2014) is particularly rapid compared with the one found by the statistics of the absorbers, (e.g. Songaila & Cowie (2010)). For example, Songaila & Cowie (2010) find a number density of LLSs per unit redshift $n(z) \propto (1 + z)^{-1.94}$, corresponding to a mean free path evolution of MFP(z) $\propto (1+z)^{-4.44}$. For comparison, Crighton et al. (2019) measure a redshift evolution for the incidence of LLS of $l(z) \propto (1 + z)^{1.70}$. These differences however are not surprising, since different absorber populations (e.g. partial Lyman limit systems, Lyman limit systems, damped systems) could contribute in a different way to the mean free path and scale differently with redshift (Prochaska et al. 2010; Inoue et al. 2014).

The mean free path of HI ionizing photons is a physical quantity averaged over a sample of QSOs at high-z. It is a measure of the distribution of Lyman limit (LL) opacity and, hence, of the IGM ionization state in the Lyman continuum. More information is carried out by the probability distribution function of the individual free paths of the population, PDF(IFP). This latter depends on the distribution of neutral patches, sparse HI residuals surrounded by a widespread, ionized gas. They are constituted both by diffuse gas clouds in the IGM and/or by denser large scale structures, i.e. the external regions of galaxies or around filaments, as proposed by Fumagalli et al. (2011); Worseck et al. (2014).

The study of PDF(IFP) is important as a benchmark for numerical simulations of the IGM. Strong constraints on cosmological simulations can be provided by a detailed knowledge of the IFP distribution across a wide redshift range. For example, radiation-hydrodynamical simulations by Rahmati & Schaye (2018) indicate that, close to the reionization epoch (at $z = 6$), the distribution of IFP peaks at ~5 pMpc with an extended tail towards smaller values, if measured along a random line of sight. If instead the line of sight is centered on massive galaxies, they predict a bimodal distribution with a peak at few kpc and another, less pronounced, peak at ~5 pMpc.

In this paper we aim at studying the distribution function of the HI ionizing individual free paths centered on relatively bright ($M_{1450} \lesssim -26$) QSOs at $3.6 \leq z \leq 4.6$ and try to understand the connection between the IFP and the LyC escape fraction. This work can be useful to understand the physical conditions that allow HI ionizing photons to escape from QSOs and subsequently ionize the IGM.

The paper is organized as follows: in Section 2 we describe the SDSS data used in this work, in Section 3 we outline the adopted method to measure the QSO physical properties (i.e. spectral slope, Lyman Continuum escape fraction, individual free path of HI ionizing photons). Section 4 describes the results, Section 5 investigates the connection between f_{esc}(LyC) and free path. Discussions, summary, and conclusions are provided in Sections 6 and 7, respectively. In Appendix A we provide an example of SDSS QSO spectrum, while in Appendix B we discuss the difference between the MFP obtained from stacked spectra and from the mean value of the IFP distribution.

Throughout the paper, we assume $H_0 = 70$ km/s/Mpc, $\Omega_m = 0.3$, and $\Omega_\Lambda = 0.7$. Physical distances are expressed in proper Mpc (pMpc).

2. Data

2.1. The SDSS DR14 QSO sample

The 14th Data Release of the Sloan Digital Sky Survey (SDSS, DR14) contains observations of different sources **till** July 2016, including all the previous detections in the precedent phases for a total of 526.356 QSOs detected over 9376 deg^2 (Paris et al. 2018). These sources have been selected combining many color criteria at different redshifts from several surveys; in particular, to determine the redshift and the type of a source, the automated procedure of the SDSS makes a comparison between the observed spectra and a set of synthetic ones. At each supposed redshift, a least-squares fit to each observed spectrum is done, using a general set of models for galaxies, QSOs, and cataclysmic variables. For each extragalactic model, a range of redshifts is explored and a chi-squared analysis is computed; the best five templates of each category are then stored in order of increasing reduced chi-squared. Finally, these latter models are refitted enhancing the resolution on the spectra (diminishing the initial velocity step); in this way, obtained the overall best fit, the class and redshift of each source are specified (for a more accurate description of the targets spectral classification and redshift measurement made by SDSS, see Bolton et al. (2012)). The spectroscopic data of the QSOs have been taken in the wavelength range $3600 \lesssim \lambda_{obs} \lesssim 10000$ Å, with spectral resolution varying from ~ 1300 at 3600 Å up to ~ 2500 at 10000 Å.

We have selected a sample of bright QSOs with $17.0 \leq I \leq 20.0$ ($-29.0 \lesssim M_{1450} \lesssim -26.0$) and $3.6 \leq z \leq 4.6$ from DR14 for a total of 2840 objects, analyzing their UV/optical spectra in order to characterize their physical properties and the surrounding IGM. The relative wide ranges in magnitude and redshift allowed us to investigate a possible luminosity and redshift evolution of our targets. In particular, the choice of this redshift coverage is mainly due to the following reasons: 1-QSOs at redshift $3.0 < z < 3.6$ present a bias against having $(u - g) < 1.5$ for which sources with strong Lyman limit absorptions are mostly selected with respect to the other ones, altering the estimation of their physical properties in this redshift range (Prochaska et al. 2009). 2-At this epoch of the Universe, the transmission of the IGM is still relatively high, allowing the observations of the QSO characteristic features.

3. The Method

3.1. Cleaning of the sample

In order to proceed for a statistical analysis on the characterizing properties of our QSO sample, each spectrum has been individually explored, and several fundamental information extracted (i.e. the spectral slope, the free path of UV ionizing photons, the Lyman continuum escape fraction).

At first, a visual inspection of all the $3.6 \leq z \leq 4.6$ QSOs has been undertaken, so to obtain a least biased sample as possible; in this way, we have checked for the type of source and its redshift, automatically assigned by the SDSS pipeline. In fact, several targets have been wrongly classified as QSOs by the SDSS procedure, although they do not present any typical feature of a QSO UV/optical spectrum at these redshifts; for this reason, these spurious sources have been discarded from the final sample. Furthermore, a precise determination of the systemic redshifts of the sources is of great importance in order to reduce the uncertainties on the estimation of the parameters describing the QSOs; therefore, each spectrum has been checked for inaccurate

position of the strongest lines in emission of the QSO. To eventually refine the spectroscopic redshift of a source, the OI emission line (1305.5 Å, rest-frame) has been used; this line forms in the outermost portion of the Broad Line Region (BLR) and is thus less affected by changing in its spectral shape due to possible QSO outflows (Matsuoka et al. 2005). Taking the wavelength relative to the OI line centroid (λ_{OI}, in Angstrom), the new redshift has been obtained as $z_{new} = \lambda_{OI}/1305.5 - 1$. Moreover, the modified redshifts of a few objects turned out to fall outside of the redshift interval used to select our QSOs. Excluding these latter sources and the spurious ones, the final sample consists of 2508 QSOs, each of which has been analyzed in order to obtain information about their evolution and the environment in which they reside.

3.2. Estimating the spectral slope

The spectral slope of each QSO has been obtained by fitting the region red-ward of the Lyα line ($\sim$1216 Å rest-frame) with the power law $F_\lambda = F_0 \lambda^{-\alpha_\lambda}$, and then extrapolating it in the bluer side of the spectra, imposing a softening at shorter wavelengths with $\alpha_\lambda^p = \alpha_\lambda - 0.72$ at $\lambda_{rest} \lesssim 1000$ Å (Cristiani et al. 2016; Stevans et al. 2014; Shull et al. 2012; Telfer et al. 2002).

At first, we have corrected each spectrum for interstellar extinction using the Cardelli et al. (1989) extinction law and the absorptions available in the SDSS database. Then, following Cristiani et al. (2016), the five spectral windows free of emission lines, reported in their Table 1, have been used for the fit. In particular, to estimate the intrinsic continuum, the average values (both in flux and wavelength) obtained in each window, after an iterative 2σ clipping, have been exploited. Leaving the normalization constant F_0 and the spectral slope α_λ as free parameters, a non-linear least-squares minimization routine has been used in order to obtain the best-fit parameters which described the local continuum of each QSO. Extrapolating the power law in the region blue-ward of the Lyα line, the flux decrements DA and DB (Oke & Korycansky 1982) have also been obtained, in the spectral windows between the Lyα and Lyβ lines (1170 - 1050 Å, rest-frame) and between the Lyβ line and the Lyman limit (1015 - 920 Å, rest-frame), respectively. **We have derived the flux decrements DA and DB in order to check the possible dependencies of the escape fraction and of the free ionizing path from these (and other) physical parameters.** As an example, Fig. 1 shows the results of this procedure applied on the **QSO SDSS J105340.8+010335.7** at $z_{spec} = 3.66983$. Here, the cyan regions are the five spectral windows used to fit the continuum, where the third window (1440 - 1465 Å, rest-frame) is zoomed to show the effect of the iterative 2σ clipping. Moreover, in several cases, the first window (1284 - 1291 Å, rest-frame) could be affected by the emission of the NV line, therefore the fit with and without this region has been computed (orange and green solid curves, respectively). **The longest wavelength region, instead, could be affected by poor sky subtraction problems, especially at high-z.** However, by default, the spectral slope for each QSO has been estimated considering the fit over all the spectral windows, in order to rely on all the exploitable data; in the upper left (right) side of Fig. 1 the values obtained for the all windows (no first window) case are shown. Finally, the yellow portions of the spectrum represent the regions in which the flux decrements DA and DB have been calculated.

Absorption features and/or noise in the spectra could even alter the slope of several sources, leading to an incorrect estimation of the intrinsic continuum of many QSOs. To avoid this kind
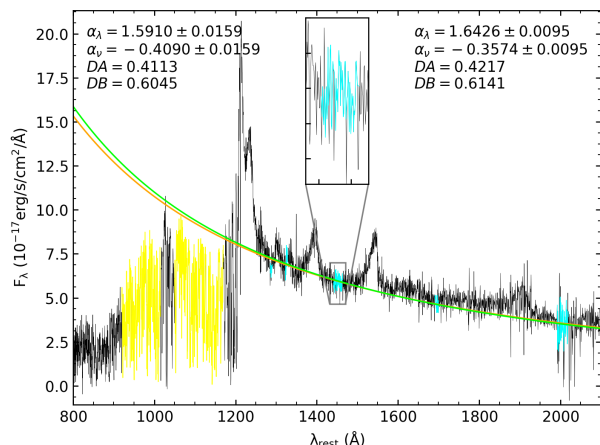
Fig. 1. Spectral slope and flux decrements for the **QSO SDSS J105340.8+010335.7** at *z* = 3.66983. The cyan portions of the spectrum are the five windows free of emission lines. The yellow regions are the places of the spectrum in which the flux decrements DA and DB have been calculated. The solid orange and green lines represent the fit with and without the first window (1284 - 1291 Å rest-frame), respectively.

of issue, the sources have been inspected one by one, removing the spectral windows affected by strong noise and recomputing the fit to estimate the spectral slope.

3.3. Estimating the Lyman Continuum escape fraction of SDSS QSOs at *z* ~ 4

The Lyman Continuum escape fraction for each QSO of our sample has been derived following Grazian et al. (2018). In summary, we estimate the mean flux density short-ward, $F_\nu(900)$, and long-ward, $F_\nu(930)$, of the Lyman limit (912 Å rest-frame) and measure the escape fraction as $f_{esc} = F_\nu(900)/F_\nu(930)$.

More precisely, the $F_\nu(900)$ is the mean flux density of the QSO in the HI ionizing continuum region, between 892 and 905 Å rest-frame, while $F_\nu(930)$ is the average flux density in the non-ionizing region red-ward of the Lyman limit, between 915 and 960 Å rest-frame. It is worth noting that this latter spectral window has been slightly enlarged with respect to the one used by Grazian et al. (2018), i.e. 915-945 Å. **In fact, after several tests, we have checked that with this new choice, the measurement of the mean flux density $F_\nu(930)$ for our sources has turned out to be more stable than the previous one, being less affected by possible absorption lines in the Lyman-α forest.** However, on average the old and new escape fractions are quite similar to each other. Finally, in calculating $F_\nu(930)$, the region between 935 and 940 Å has been avoided, due to the Lyman-ϵ emission line of the QSO. At *z* ~ 3.6−4.6, the adopted rest-frame wavelengths 892-960 Å correspond to observed wavelengths of λ_{obs} ~ 4100 − 5400 Å, which are well sampled by the SDSS spectra we used in the present paper. In order to avoid the presence of intervening absorbers in these regions, both $F_\nu(900)$ and $F_\nu(930)$ have been derived with an iterative 2σ clipping procedure.

It is worth noting here that we do not attempt to reconstruct the intrinsic QSO continuum, but we measure the observed mean flux level in these two spectral windows, assuming an equal residual Lyman series absorption above and below the Lyman

limit. The observed distribution of the escape fraction could be different from the intrinsic one due to the errors in the estimation of the local continua, both red-ward and blue-ward of the Lyman continuum. At this aim, we have carried out an estimation of the flux variation in the Lyman forest. Precisely, for each QSO studied in this paper, we measure the mean flux (after the 2σ clipping procedure) in the rest-frame wavelength 915-960 Å and 960-1005 Å and compute their flux ratio. It turns out that this flux ratio does not depend on the absolute magnitudes or spectroscopic redshifts, and its scatter is ~0.15 dex. We assume that the same error can be applied when we estimate the individual QSO continua (both red-ward and blue-ward of the Lyman continuum). Just for comparison, Prochaska et al. (2010) estimated a 10% uncertainty on the estimation of the SDSS QSO continuum in the Lyman forest. Adopting this ~0.15 dex continuum uncertainty, we find that the typical scatter of the escape fraction is ~15-17% for $f_{esc} > 75\%$, while it is progressively decreasing to 1-2% at $f_{esc} < 20\%$. Moreover, the output values of the escape fraction are fully consistent with the input values, without any systematic. We have checked that the mean value of the escape fraction distribution is not perturbed significantly by the scatter due to flux ratio uncertainties. Interestingly, the intrinsic distribution (i.e. after deconvolution by the observed scatter) gives slightly higher mean values of f_{esc} than the observed distribution. In the following, given that the (reconstructed) intrinsic distribution is too noisy, we prefer to plot only the observed distribution.

As discussed in Grazian et al. (2018), this technique is equivalent to measure $f_{esc} = \exp(-\tau_{LL})$, where τ_{LL} is the average opacity at the Lyman limit. It could include the cumulative effect of small absorbers ($\tau_{LL} < 2$), which cannot be individually detected in the SDSS QSO spectra (Prochaska et al. 2010).

We avoid using the spectral window close to the 912 Å rest-frame of the QSO (i.e. 905-912 Å rest-frame) since this region can be affected by the QSO proximity effect. Moreover, the uncertainties in spectroscopic redshift determination of SDSS spectra, typically of σ_z ~ 0.01, does not allow us to locate precisely the position of the Lyman limit, so we decide to place our spectral window far from this region.

It is worth noting here that our estimate of the LyC escape fraction f_{esc} for the SDSS QSOs at *z* ~ 4 can be considered a lower limit to the real value for several reasons: 1-we do not correct this value for the spectral slope of the QSOs. In the $\lambda_{rest} \leq 1200$ Å region, a typical value for the spectral slope is α_ν ~ −1.7 (Lusso et al. 2015). Correcting for this slope, the true values of f_{esc} for our QSOs should be revised upward by ~ 6%. 2-There are possible Lyman-α absorbers which serendipitously fall between the two spectral windows adopted to estimate f_{esc}, i.e. 905-915 Å in the QSO frame. Following Inoue et al. (2014), we compute a flux decrement of ~ 0.6 between 930 and 900 Å rest-frame, due to intervening IGM (see their Fig. 4 for z_{spec} = 4.0). 3-We exclude from the escape fraction calculations the proximity region of the QSO, which is instead a clear signature of the ionization of the surrounding IGM. **4-As we will discuss in Section 5.1, if we estimate the mean flux density $F_\nu(900)$ in a shorter spectral window between 898 and 905 Å rest-frame, the escape fraction turns out to be slightly larger than the one computed in the wider region between 892 and 905 Å rest-frame.**
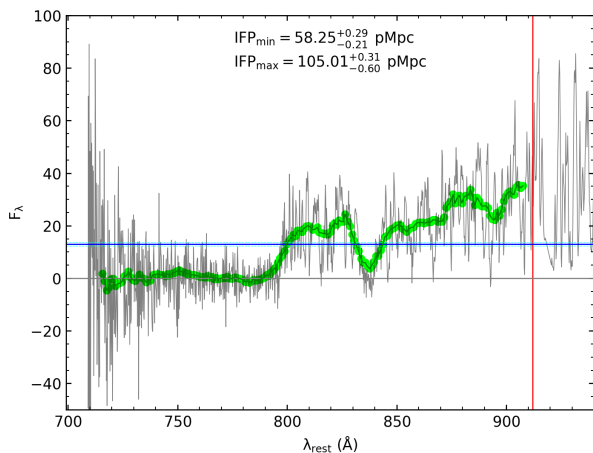
Fig. 2. Individual free path estimation for the **QSO SDSS J114514.2+394715.9** at z_{QSO} = 4.06557. The green circles and solid line represent the average fluxes in the 10 Å windows and the spline, respectively. The red line marks the Lyman limit. The horizontal blue and cyan lines represent f_{red} and its uncertainty, respectively. **In this case, the choice of IFP corresponds to $\lambda_{rest} \sim 800$ Å.** The flux on the y axis is in arbitrary unit.

3.4. Estimating the Individual Free Path of HI ionizing photons

The free path of photons at 912 Å rest-frame is a fundamental parameter in order to estimate the HI photo-ionization rate. The IFP is defined as the physical distance a 912 Å rest-frame photon can travel before being attenuated by a factor of e^{-1} in flux. Such an attenuation can be caused by encountering a Lyman limit system with optical depth $\tau_{HI} = 1$ or, alternatively, by the cumulative effect of a great number of more transparent absorbers, with $\tau_{HI} << 1$, even if it is not possible to detect them individually.

In order to carry out a statistical analysis, the IFP has been estimated individually for each spectrum. At first, the SED of each source was corrected for its intrinsic spectral slope, considering a break in the continuum at ~ 1000 Å rest-frame, as found by Stevans et al. (2014); Shull et al. (2012). This has been done flattening the spectral slope of each QSO (obtained as described in Section 3.2) for the value $\Delta\alpha_\nu = 0.72$ (Cristiani et al. 2016) at $\lambda \leq 1000$ Å rest-frame. Starting from $\lambda_{in} = 3600/(1 + z)$, the region blue-ward of the Lyman limit has been divided into several spectral windows of width $\Delta\lambda$ and upper limit $\lambda_{fin} = \lambda_{in} + \Delta\lambda$. By default, $\Delta\lambda$ has been chosen to be 10 Å, being this value a good compromise in following the spectral features of the specific QSO without under/overestimating the IFP; in this way, the spectral window has been moved upward of 1 Å in every cycle until the Lyman limit, i.e. $\lambda_{fin} = 912$ Å, has been reached. In each window, the average fluxes and wavelengths have been calculated through an iterative 2σ clipping on the respective data, and then interpolated with a spline. In this case, an asymmetric clipping has been computed in order not to underestimate the continuum level of the spectrum: all the negative residuals have been rejected. The average flux obtained in the upper window near the Lyman limit, f_{912}, has been used to compute the flux reduced by a factor e^{-1}, as $f_{red} = f_{912}/e$. The intersection between the spline and f_{red} defines the wavelength relative to the (Lyman limit) absorber along the line of sight, λ_{LLS}, whose redshift has been obtained as $z_{LLS} = \lambda_{LLS}/912 - 1$. Therefore, the comoving
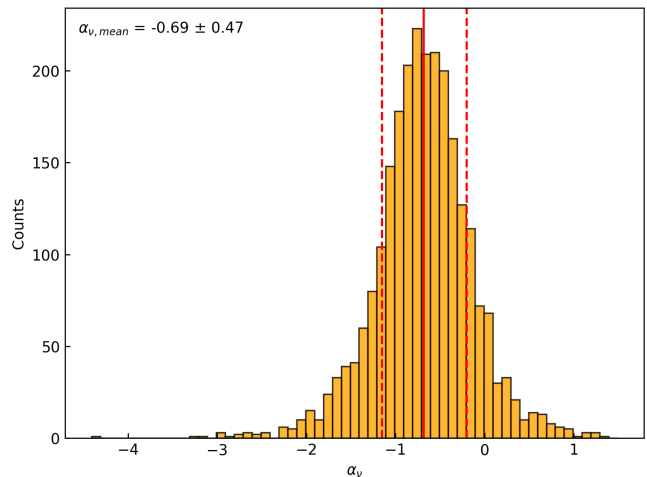
Fig. 3. Spectral slope distribution. The vertical red solid and dashed lines are the mean value and its uncertainties, respectively.

distances between the absorber and the source have been computed as $D_c = D_c(z_{QSO}) - D_c(z_{LLS})$, where z_{QSO} is the redshift of the QSO. Finally, the free path of ionizing photons is the proper distance between the QSO and the LLS, and it has been derived as IFP = $D_p^{LLS} = D_c/(1 + z_{LLS})$. **We have checked that the IFP value does not depend on the choice of $\Delta\alpha_\nu = 0.72$. We have derived the individual IFPs for all the QSOs with $\Delta\alpha_\nu = 0.5$ and $\Delta\alpha_\nu = 1.0$, which are still allowed by the uncertainties provided e.g. by Stevans et al. (2014); Shull et al. (2012); Telfer et al. (2002) and found that the effect is negligible, with a difference of ~ 0.1 pMpc on the mean value of IFPs.**

In several cases, absorption features blue-ward of the Lyman limit could drag the interpolated fluxes below f_{red}, and then bring it above this threshold; this trend could create many intersections of the spline with f_{red} for each of which the individual free path has been computed. **Both the minimum and maximum IFP values have been derived, as can be seen in Fig. 2 for the QSO SDSS J114514.2+394715.9 at z_{QSO} = 4.06557. For all our QSOs, the real individual free path of ionizing photons is clearly associated to the maximum IFP value, being the minimum one just the result of an absorption feature; for this reason, the maximum IFP has been set as the definitive value for each QSO by default.**

4. Results

4.1. Spectral slope distribution

To study the ensamble properties of our QSO sample, we have at first obtained the spectral slope distribution in order to carry out a statistical analysis of this parameter and to explore possible evolution in magnitude and/or redshift.

Fig. 3 shows the distribution function with the red line marking the spectral slope mean value, $\alpha_\nu = -0.69 \pm 0.47$ (i.e. $\alpha_\lambda \approx 1.31$), and where the uncertainties (red dashed lines) have been computed as half of the difference between the 84th and 16th percentiles. This result is in agreement with previous estimations of Cristiani et al. (2016), who found $\alpha_\lambda = 1.30$ at $3.6 < z < 4.0$, and those of Telfer et al. (2002) and Stevans et al. (2014) who obtained $\alpha_\lambda \approx 1.31$ and 1.17 at lower redshift, respectively.

To investigate the evolution of the spectral slope, the whole QSO sample has been split in two halves including fainter and
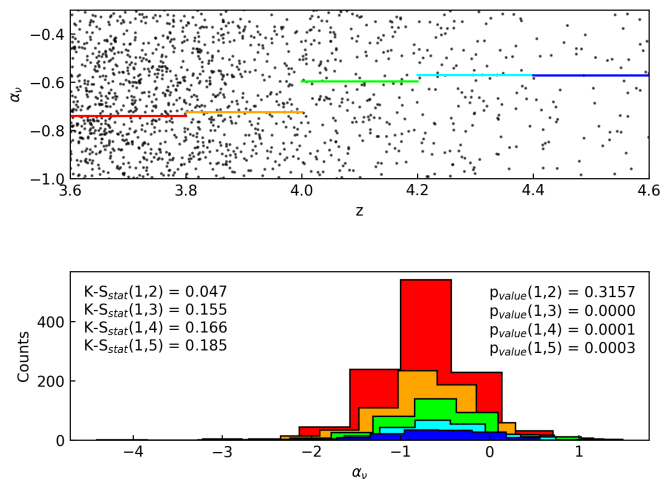
Fig. 4. *Upper panel*: spectral slope as a function of redshift. The horizontal red, orange, green, cyan, and blue lines are the average values from the lowest to the highest redshift bins. *Bottom panel*: distribution of the spectral slope in each redshift bin (with the same color of the upper panel).

Fig. 5. Lyman Continuum escape fraction distribution for the whole sample (in red) and for sources with IFP $\geq$ 10 pMpc (in blue). The solid and dashed grey lines represent the mean values of the former and latter distributions, respectively.

brighter sources with respect to different absolute magnitude thresholds M_{1450} (from ≈ -29.0 up to ≈ -26.0); the two distributions corresponding to each M_{1450} have been compared using a two-sample Kolmogorov-Smirnov (KS) test, choosing to reject the null hypothesis (i.e. the hypothesis that two indipendent samples are drawn from the same parent distribution) for a p_{value} below 1 %. With this test, a luminosity evolution of the spectral slope has been rejected.

The same procedure has been used to study the evolution of α_ν with redshift. In this case, the QSO sample has been split in five redshift bins having $\Delta z = 0.2$, in order to test possible differences among the various distribution functions, as shown in Fig. 4. The horizontal solid lines in the upper panel of the Figure represent the average values of the spectral slope in each bin of redshift, increasing from $\alpha_\nu \approx -0.74$ in the first bin up to $\alpha_\nu \approx -0.57$ in the last one. The corresponding distributions are plotted in the lower panel where the KS tests between the first and the other redshift bins are also shown. It is interesting to note that the KS tests between the first and the three last distributions are significant, defining a possible redshift evolution of the analyzed parameter.

This is probably caused by the shorter range in wavelengths used to compute the power-law fit which affects the QSO spectra at high redshift, i.e. the fact that the reddest spectral window in the spectra could exit the wavelength coverage of the SDSS for these sources. In order to test this hypothesis, we have recomputed the spectral slopes for our QSO sample excluding the spectral window at the highest wavelengths when fitting the power law. In this case we obtain an increase of α_ν from ≈ -0.65 at $z \sim 3.7$ to ≈ -0.57 at $z \sim 4.5$. Thus, at low redshift ($z \lesssim 3.9$), the average value of the slope is slightly higher than the one obtained with all the free emission windows, while it remains unchanged at high redshift. In this case, the KS test is no more significant within the different bins in redshift. For these reasons, the above-mentioned redshift trend of the spectral slope seems to be mostly caused by a systematic effect due to the reduced spectral range, rather than to a physical evolution.
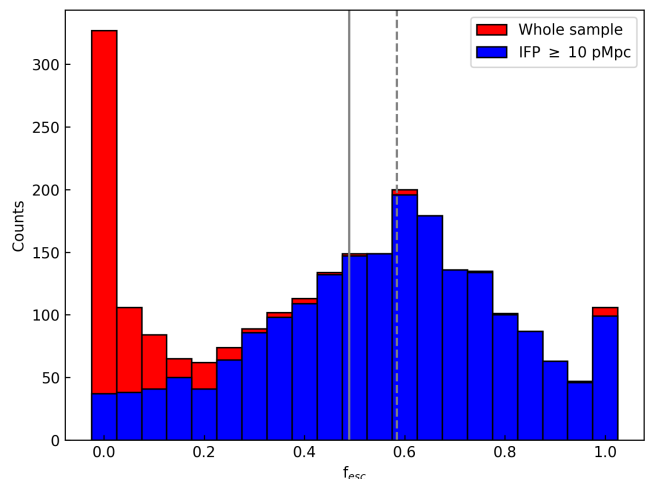
4.2. Lyman Continuum escape fraction distribution

Depending on the intrinsic properties of each specific QSO and on its possible absorption features, the computed LyC escape fractions cover a broad range of values, going from approximately zero to 100%. We have found an average value $f_{esc} = 0.49 \pm 0.36$ in our sample, as reported in Fig. 5 which shows the escape fraction distribution function of all our sources (red histogram); this one results in a multi-modal distribution with a peak at low values (**14.5% of the QSOs with $f_{esc} < 0.10$**), a minimum between $0.10 \leq f_{esc} \leq 0.3$ and a broad dispersion at higher values, as also found by Cristiani et al. (2016). Also shown in this figure is the distribution from the sub-sample excluding sources with IFP < 10 pMpc, resulting in an average value of $f_{esc} = 0.63 \pm 0.26$ (in blue, see Subsection 5.1). In particular, individually inspecting the sources within the first bin, the low f_{esc} peak has been attributed to the presence of several QSOs with LLSs which are intrinsic to the QSO itself or intervening along the line of sight, which absorb almost all the UV radiation escaping from them. This causes a rapid decline of the emitted flux in the region blue-ward of the Lyman limit producing a negligible escape fraction (see sub-section 5.2 for a discussion on the effect of these sources on the final statistical analysis of the QSO physical parameters).

As also done for the spectral slope, the dependence of the LyC escape fraction on both redshift and magnitude has been tested by splitting the whole sample in sources with lower and higher redshifts ($z \leq 4.1$ and $z > 4.1$), and QSOs brighter and fainter than a given absolute magnitude threshold, respectively. The different populations of objects have been compared with a KS test whose results have allowed us to exclude a possible evolution of f_{esc} with redshift. Regarding the dependency of the escape fraction on the magnitude, splitting the sample in different magnitude bins, we have obtained $p_{value} = 0.007$ in the brightest bin ($M_{1450} < -27.4$) which could be significant for a possible trend of the escape fraction with luminosity. However, at fainter magnitudes, p_{value} increases to values higher than 1%, not allowing to confirm an evolution of f_{esc} with luminosity.
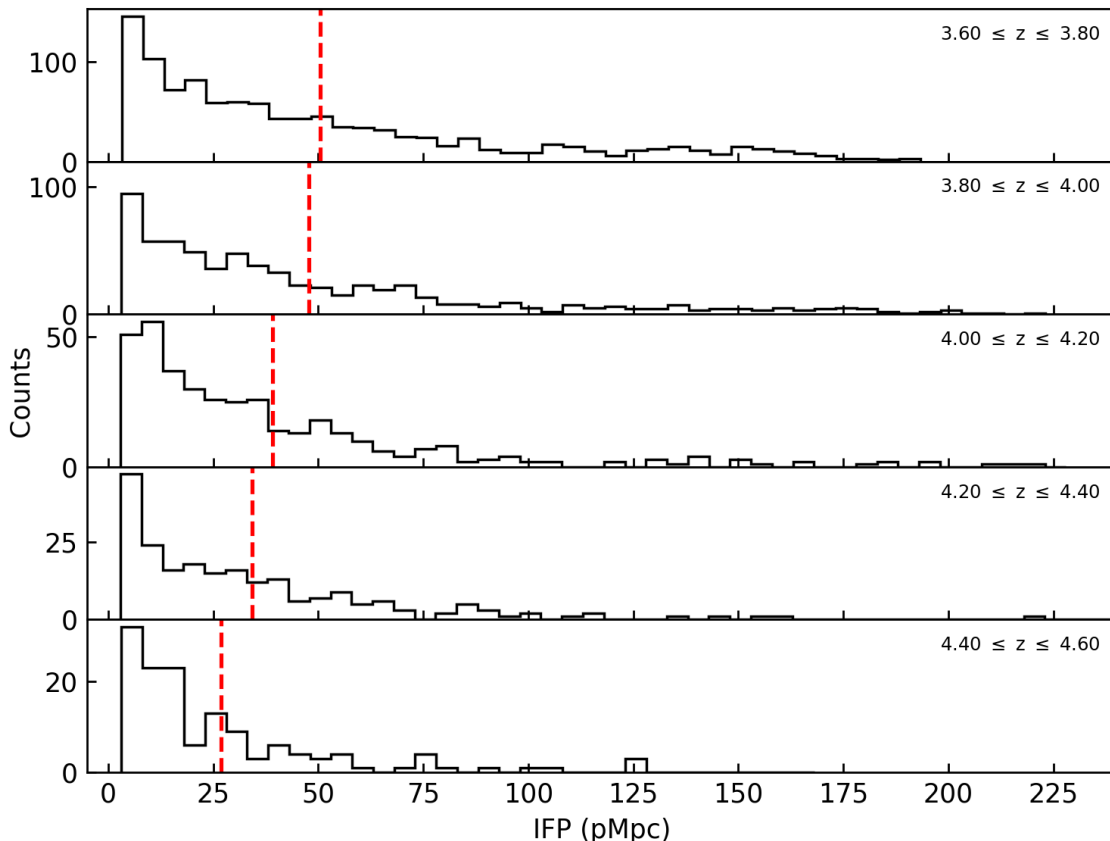
Fig. 6. Free path distribution functions for the whole QSO sample at $3.6 \leq z \leq 4.6$ in redshift bins of $\Delta z = 0.2$. The dashed red lines represent the mean value in each redshift bin.

4.3. Individual Free Path distribution

The analysis implemented in this paper has led to estimate, for the first time, the mean free path of UV ionizing photons of a QSO sample at redshift $z \sim 4$ in a statistical way, i.e. from the distribution functions of IFPs. These ones are displayed in Fig. 6 for redshift bins of $\Delta z = 0.2$, where the dashed red lines mark the mean value in each bin; a peak in the distributions is present at low free paths, with a broad tail at higher values, and a mean free path MFP ~ 43 pMpc at $z \sim 4$.

It is well known that this parameter evolves in redshift, being the Universe increasingly more and more composed of neutral hydrogen which absorbs UV photons approaching the EoR (see Sec. 5.2). At $z < 5$ the HI-ionizing background is quasi-uniform (e.g. Meiksin & White 2004; Becker et al. 2015), resulting in a skewed but uni-modal distribution of free paths that reflects Poisson variance in the incidence of (partial) Lyman limit systems along individual lines of sight. The dependency of the free paths with the luminosity of the QSOs will be discussed in the next section.

5. The Free Path versus LyC Escape Fraction Connection

5.1. Dependence of LyC escape fraction on QSO Individual Free Path

The distribution of the LyC escape fraction of our QSOs is multi-modal, with a narrow peak at $f_{esc} \lesssim 0.2$ and a broad distribution from 0.3 to 1.0, as discussed in the previous section. In order to investigate the origin of this multi-modality, we check for any dependencies of f_{esc} on the other physical parameters. No dependence has been found comparing against spectral slope α_ν, redshift, absolute magnitude M_{1450}, and Lyman decrements (DA, DB). The only exception is against the individual free paths, as shown in Fig. 7. It is interesting to note the trend of the escape fraction with the individual free path, at IFP ≤ 20 pMpc. In particular, a deficiency of high escape fractions at low IFP and a lack of low f_{esc} at high free paths are evident. Moreover, an increase of the escape fraction proportional to the individual free path seems to be present up to IFP of the order of $\sim 10 - 20$ pMpc.

About the lack of sources at low free paths and high escape fractions, this could be interpreted as an effect due to the choice of the spectral window in the ionizing region, between 892 and 905 Å rest frame, adopted in the f_{esc} derivation. Figure 8 shows the estimation of the IFP for the **QSO SDSS J034402.9-065300.6** at $z = 3.92455$. This source has IFP $= 6.7 \pm 0.3$ pMpc and $f_{esc} = 0.0$, thus consistent with the above consideration. In this case, the wavelength associated with the individual free path is inside the ionizing window of the escape fraction (highlighted in orange in the Figure), creating a correlation between the two parameters. At the same time, a reduced spectral window, closer to the Lyman limit at 912 Å, could increase the average flux in this region, leading to an enhanced escape fraction for the specific QSO. As an example, the rest-frame 892-905 Å window was reduced to 898-905 Å and the escape fractions recomputed

using this modified window for the whole sample. This is displayed in Figure 7; here, the escape fractions adopted in this paper (**grey** points, calculated as explained in Section 3.3) and the modified ones (**orange** points) are plotted for each source at $z > 3.80$ (see Section 6.1.1) as a function of their free paths. As can be seen, the **orange** points tend to populate the low IFP region at high escape fractions, reducing the previously discussed deficit of these sources. This confirms that the lack of these QSOs is mostly due to the spectral window (**892-905 Å rest-frame**) adopted for calculating the escape fraction rather than to a physical mechanism. In principle, adopting an even shorter window close to 912 Å rest-frame for the f_{esc} derivation would further reduce the number of sources with low IFP and low f_{esc} but, at the same time, the scatter would significantly increase. For these reasons, we decide to fix the spectral windows for the escape fraction measurements to the rest-frame 892-905 Å wavelength range. Interestingly, Cristiani et al. (2016) adopted a wavelength interval much further than the one adopted here (850-880 Å rest-frame), and their peak at low f_{esc} is slightly enhanced compared to our. At $f_{esc} \geq 0.20$ Cristiani et al. (2016) have a fraction of 69.5% of their sample, while we have 78.8% of the QSOs above this threshold.

In Fig. 5 we plot also the escape fraction distribution (in blue) in case the sources with short IFP have been removed. Limiting the sample to IFP ≥ 10 pMpc, the peak at $f_{esc} \sim 0$ disappears, and a continuous trend is in place, indicating that the sources with low escape fraction are probably due to absorbers falling by chance at small distance from the QSO, or associated to it. In the following sections we discuss about the nature of these absorbers.

Another interesting feature emerges from Fig. 7: excluding very few objects with $f_{esc} \sim 20\%$ and large IFP, a trend of larger escape fraction appears for longer IFP, as highlighted by the mean of f_{esc} in IFP bins of 30 pMpc (black filled and open square) and by the lower envelope of the individual QSOs (grey points). In particular, at very large IFP (> 170 pMpc), no QSO has $f_{esc} \leq 40\%$. We have checked that this correlation does not depend on the choice of the parameter $\Delta\alpha_\nu = 0.72$ adopted. The observed trend remains if we change this parameter to 0.5 or 1.0. This correlation, however, could be limited only to few QSOs with IFP greater than 170 pMpc, while for the bulk of the QSOs with $30 < $ IFP $ < 150$ pMpc the escape fraction shows no strong dependency on the IFP, remaining almost constant at $\sim 60-70\%$. Finally, we have also computed a KS test to check whether the sample of QSOs with IFP > 150 pMpc and the one with $30 < $ IFP $ < 150$ pMpc may be drawn by the same parent population. We find a p_{value} of 0.064, thus we can conclude that the differences between the two samples are not significant. At present, however, we cannot exclude that this trend is due to some systematic effects of the data (e.g. due to the color selection of $z \sim 4$ QSOs). More checks are needed in the future to substantiate such statement.

In the following sections, when dealing with the IGM properties, we will consider two extreme options to interpret the nature of the QSOs with IFP ≤ 10 pMpc: 1-all the sources with IFP ≤ 10 pMpc are affected by intervening absorbers (option 1). 2-all these QSOs are instead related to associated (intrinsic) absorption (option 2), due to the presence of dense neutral absorbers in their vicinity. In the former, the IFP is the physical distance between the emitter and the absorbing system, while in the latter the IFP reflects the outflow velocity of material which sits close to the QSO.
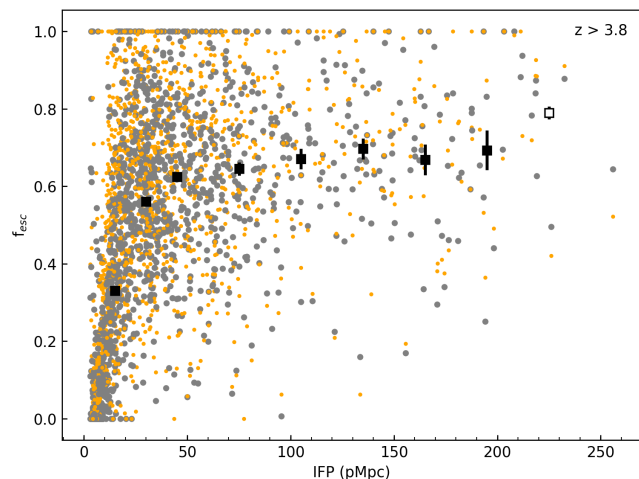
Fig. 7. Escape fraction vs individual free path for QSOs at $z > 3.80$. The grey points show the escape fractions calculated with the ionizing window 892-905 Å rest-frame; the orange points represent the escape fractions with the reduced ionizing window 898-905 Å rest-frame. The black filled square represent the mean f_{esc} in IFP bins of 30 pMpc (the open square is the mean of f_{esc} in IFP bin containing less than 10 sources); the vertical bars are the error on the mean in each bin of IFP. QSOs with $z < 3.8$ have been excluded here to avoid possible biases due to their IFP under-estimation, as discussed in Section 6.1.1.
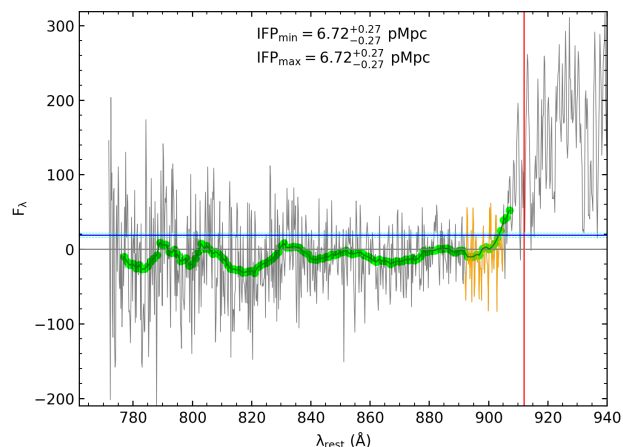
Fig. 8. Free path estimation for the **QSO SDSS J034402.9-065300.6** at $z = 3.92455$, a QSO with short IFP and low value of escape fraction, thus representing an example of correlation between IFP and f_{esc}. The orange spectral region marks the ionizing window used to calculate the escape fraction (i.e. 892-905 Å rest-frame).

In "option 1", the QSOs with IFP ≤ 10 pMpc have been included in the statistical analysis of the probability distribution function for the free paths. In "option 2", instead, they have been excluded not representing the mean properties of the IGM, since the central engines are affected by their surrounding Circum Galactic Medium (CGM) or host galaxy. There could also be associated absorbers with optical depth $\tau \leq 1$, which would mildly suppress the flux blue-ward of the Lyman limit, thus mimicking intervening absorbers with IFP ≥ 10 pMpc. We assume here that all the systems with IFP ≥ 10 pMpc are intervening absorbers.

This assumption will reduce the observed IFP w.r.t. the intrinsic one.

In Section 6.2 we discuss in detail the nature of the close absorbers with IFP ≤ 10 pMpc.

5.2. The redshift evolution of the Mean Free Path

Since our data cover a relatively wide redshift range, we investigate here the evolution of the mean free path of ionizing photons with the cosmic age. An evolution in redshift of the mean free path is certainly expected, due to the properties of the IGM and its evolution with the cosmic expansion. It is indeed known that this parameter should evolve in redshift, being the Universe increasingly composed by neutral hydrogen which absorbs UV photons approaching the EoR. A precise estimation of the redshift evolution of MFP is fundamental, since it is one of the primary sources of uncertainty in the calculation of the photoionization rate Γ_{HI}, as we will discuss in the next sub-section.

In the past, the mean free path of ionizing photons has been estimated mainly by the number statistics of LLSs with redshift (e.g. Songaila & Cowie 2010; Prochaska et al. 2013; Inoue & Iwata 2008; Inoue et al. 2014) or by stacking a large number of QSO spectra at the same redshifts (e.g. Prochaska et al. 2009; Worseck et al. 2014). Given the high quality of the SDSS spectra analyzed here for the escape fraction determination, we decided to measure the free path for each individual QSO at $3.6 \leq z \leq 4.6$ in order to derive the mean value and scatter of this parameter, leading to a statistical analysis of the free path distribution. Moreover, the PDF of IFP gives better complementary information than a single average value (MFP).

At the beginning, we compute the individual free paths for SDSS QSOs at $3.6 \leq z \leq 4.2$ and $I \leq 19.5$ and find that the mean values of the IFP distribution are larger than the MFP obtained by Prochaska et al. (2009) and Worseck et al. (2014) through stacking technique. We then extend the magnitude limit of our QSO sample to $I = 20.0$ and derive mean values of IFP which are consistent with the Prochaska et al. (2009) and Worseck et al. (2014) results. We then check the dependence of the IFP from the absolute magnitude M_{1450} of the SDSS QSOs, since we have indications of a possible trend of the IFP with QSO luminosity.

Fig. 9 shows the dependence of the IFP from the absolute magnitude M_{1450} of SDSS QSOs at $3.8 < z < 4.0$ and $I \leq 20.0$. The small black dots represent the measurements for individual QSOs, while the big blue circles indicate the mean values of the free paths and absolute magnitudes when dividing the sample in bins of ~ 1.0 in M_{1450}. A trend of larger IFP for brighter sources is possibly suggested by this plot, with IFP ~ 60 pMpc at $M_{1450} \sim -28$ and IFP ~ 45 pMpc at $M_{1450} \sim -26$. This trend is probably due to sources with low IFP which are more frequent at fainter magnitudes (in the lower right corner of Fig. 9), rather than the lack of faint QSOs with long free paths.

A similar trend has been found also at higher redshifts, at least up to $z = 4.6$, the maximum redshift for our sample. Fig. 10 shows the mean values of the free paths for all the SDSS QSOs with $3.6 \leq z \leq 4.6$ and $I \leq 20.0$, divided in redshift and absolute magnitude M_{1450} bins. The continuous lines are the best fit with a linear relation to the mean values. The trend with luminosity becomes milder going at higher redshifts, and at $z \sim 4.5$ is almost negligible.

While a redshift evolution of the mean free path is expected both by theoretical and observational arguments, as discussed above, however, an evolution in magnitude of the MFP is not so obvious. In principle, we should expect that the physical properties of the IGM (e.g. the MFP) will be independent from the
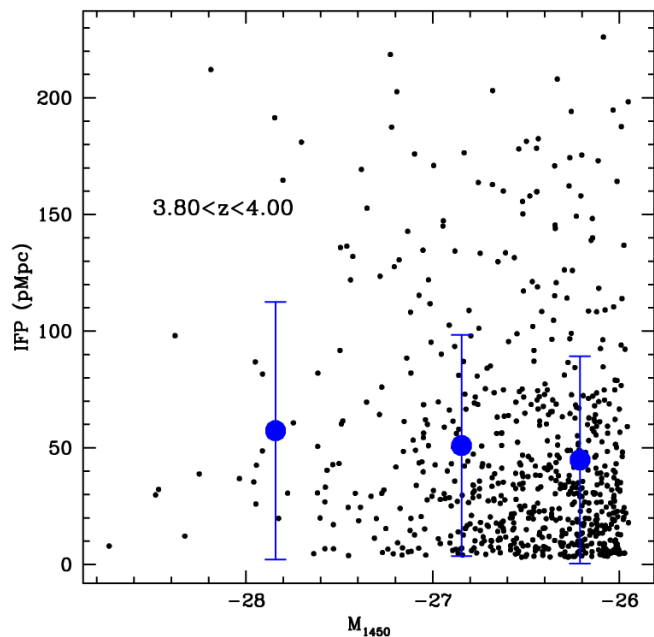

Fig. 9. The small black dots represent the free path vs absolute magnitude M_{1450} for individual SDSS QSOs at $3.8 < z < 4.0$ and $I \leq 20.0$. The big blue points show the average value of the free paths in bin of $\Delta M_{1450} = 1.0$. The error bars of the big blue points show the 84th and 16th percentiles of the individual free path distributions for each bin of absolute magnitude.

properties of the tools adopted to study it, i.e. the high-z QSOs analyzed in this paper. We have carried out a simple simulation in order to check whether the dependence of the IFP from the QSO luminosity can be due instead to an instrumental effect, e.g. to the decreasing signal-to-noise ratio of the spectra. We started from the observed spectra of five SDSS QSOs at $I \leq 17.5$, i.e. with high S/N ratio. We select them in order to span as much as possible the observed range of IFPs, from 20 to 200 pMpc. These spectra have relatively high S/N and we can measure their IFP with very high accuracy. We used these spectra as input to our simulations: for each simulated I-band magnitude in the range $I = 18 - 20$, we rescale the input spectra by the factor $10^{0.4(I_{obs} - I_{sim})}$, where I_{obs} and I_{sim} are the observed and simulated I-band magnitudes, respectively. Then we add random noise considering the original (i.e. without any rescaling) r.m.s. of the observed spectra released by SDSS and derive the IFP of the simulated QSOs with the same tools used for our measurements, as described above. We repeat this operation for 100 times for each simulated magnitude. In Fig. 11 we plot the mean values of the output IFP for each of the five QSOs and for each simulated magnitude in the range $I_{sim} = 18 - 20$, corresponding to the magnitude range of our sample. The error bars indicate the 16th and 84th percentiles of the IFP distribution. We find no dependence of the resulting IFP from the luminosity of the objects. This reassures us that the observed trend with luminosity shown in Fig. 9 is not spurious and cannot come from the low S/N of the data at $M_{1450} \sim -26$. Only at very large IFP (~ 200 pMpc) and at $I \sim 20$ there is a slightly underestimation of the free path, of the order of ~ 0.1 dex. This effect however is too small to explain the strong trend found in Fig. 9. Moreover, it is not expected that at faint observed magnitudes the SDSS algorithms are selecting preferentially sources with lower free paths w.r.t. bright QSOs. It
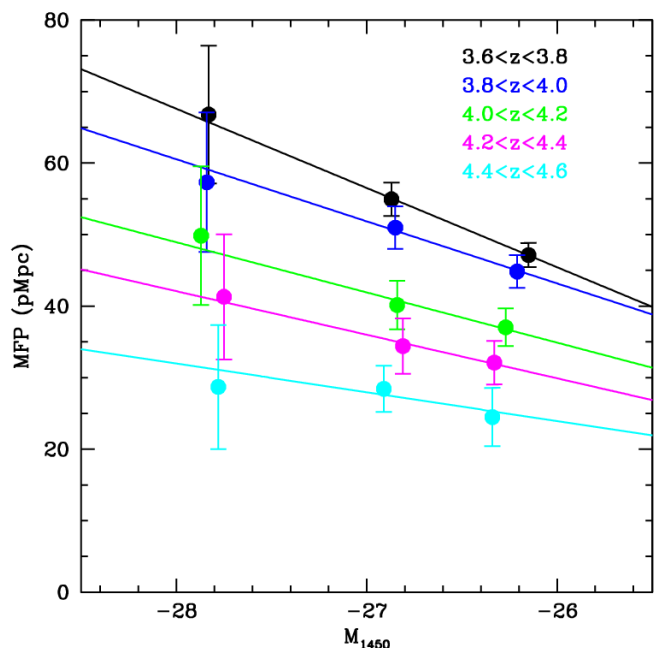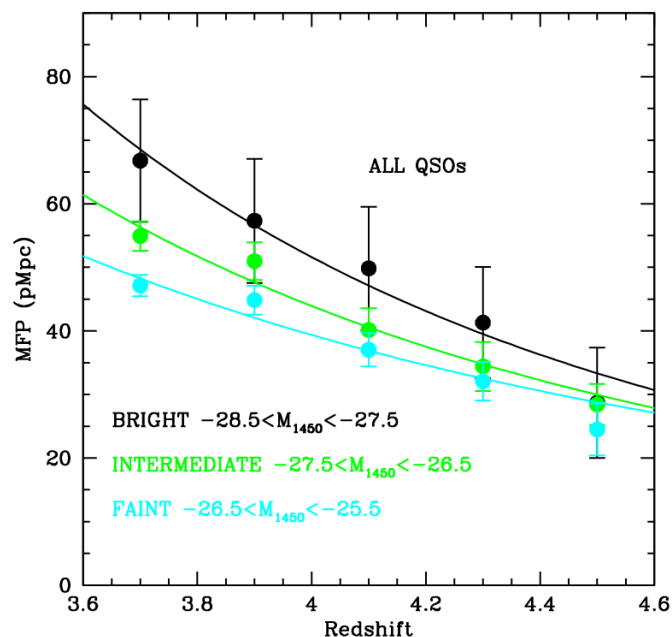
Fig. 10. The average values of the free paths for all SDSS QSOs at $3.6 \leq z \leq 4.6$ and $I \leq 20.0$, divided in redshift and absolute magnitude M_{1450} bins. The vertical bars are the standard error of the mean in each bin of absolute magnitude. The lines show the best linear fit to the observed mean points for each redshift bin.
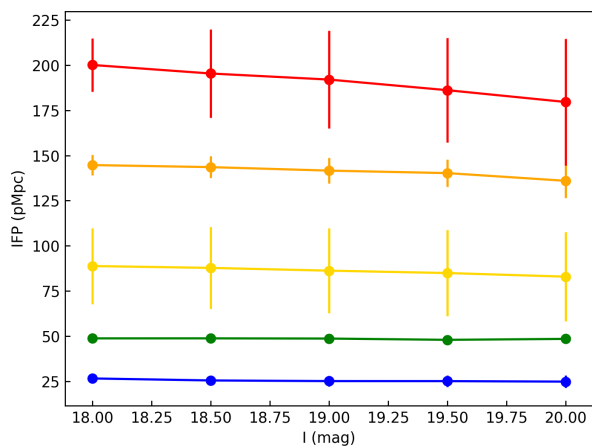
Fig. 11. The free path of simulated QSOs at different I-band magnitudes. Circles show the mean value of 100 simulations, for each QSO and magnitude bin, while the error bars indicate the 16th and 84th percentiles of the resulting IFP distribution. No strong trend of the IFP with the observed I-band magnitude has been detected.

is not so obvious to find a plausible explanation for the observed trend with luminosity.

Fig. 12 summarizes the mean values of the IFP for three different bins of absolute magnitudes of SDSS QSOs from $z = 3.6$ to $z = 4.6$ and magnitude $I \leq 20.0$. The mean values of IFP have been fitted with a power-law relation $\mathrm{MFP}(z) = \mathrm{MFP}(z = 4) \times [(1 + z)/5]^{\eta}$ and the continuous lines show the best fit at different luminosities. The redshift evolution η we are finding is milder than the one by Worseck et al. (2014). **If we exclude from the fit the redshift interval between $z = 3.6$ and 3.8 (we**

Fig. 12. The mean values of the IFP for SDSS QSOs at $3.6 \leq z \leq 4.6$ and $I \leq 20.0$ ("option 1"), divided in redshift for three bins of absolute magnitude M_{1450}. The vertical bars are the standard error of the mean in each bin of absolute magnitude. The continuous lines show the best power-law fit to the observed mean points for each M_{1450} bin.

Table 1. Redshift evolution of the MFP for all QSOs at $3.6 \leq z \leq 4.6$ and $I \leq 20.0$ ("option 1"). The first two columns are the best fit slope and normalization at $z = 4$, respectively. The last column indicates the absolute magnitude bins. For comparison, the best fit values by Worseck et al. (2014) are $\mathrm{MFP}(z = 4) = 37 \pm 2$ pMpc and $\eta = -5.4 \pm 0.4$. In the bottom part, we have excluded the $3.6 < z < 3.8$ redshift bin from the fit of the $\mathrm{MFP}(z)$ relation, since it could be affected by a possible underestimation of the free path due to the limited wavelength coverage of the SDSS spectra in the bluer wavelengths.

η	$\mathrm{MFP}(z = 4)$ pMpc	Magnitude bin M_{1450}
	All redshift bins	
-4.6 ± 0.6	51.6 ± 1.6	$-28.5 \leq M_{1450} \leq -27.5$
-4.0 ± 0.5	43.9 ± 1.1	$-27.5 \leq M_{1450} \leq -26.5$
-3.3 ± 0.6	39.4 ± 1.2	$-26.5 \leq M_{1450} \leq -25.5$
	Excluding 3.6 - 3.8 redshift bin	
-5.3 ± 1.0	53.2 ± 2.4	$-28.5 \leq M_{1450} \leq -27.5$
-5.1 ± 0.3	45.7 ± 0.5	$-27.5 \leq M_{1450} \leq -26.5$
-4.7 ± 0.4	40.8 ± 0.6	$-26.5 \leq M_{1450} \leq -25.5$

will see in the next sections that this bin could be prone to a systematic effect), the resulting η is consistent with the one by Worseck et al. (2014), as shown in the bottom part of Table 1. It should be noted also that the redshift range studied by Worseck et al. (2014) is $2.3 < z < 5.5$, **while our redshift interval is smaller** ($3.6 < z < 4.6$) **and thus could be more prone to uncertainties in the derivation of η.**

Interestingly, from Table 1 it is shown that, for our fainter absolute magnitude bin $-26.5 \leq M_{1450} \leq -25.5$, the best fit to our mean values has a normalization $\mathrm{MFP}(z = 4) = 39.4$ pMpc, which is quite close to the best fit value of 37 pMpc provided by Worseck et al. (2014). This is not surprising, given the fact that their results are expected to be dominated by faint QSOs with

$M_{1450} \sim -26$, which is close to the completeness limit of SDSS survey at $z \sim 4$. **The MFPs measured both by Prochaska et al. (2009) and by Worseck et al. (2014) are based indeed on a stack of 150 QSOs for each redshift bin adopted, with a bootstrap on the whole SDSS sample. Since these 150 QSOs are chosen randomly from the whole SDSS QSO survey in the adopted redshift bin, the sample is possibly dominated by fainter QSOs, and thus with lower IFPs, if the observed IFP-Luminosity relation turns out to be correct. In order to deepen the observed correlation between the IFP and the luminosity, we have cut our sample to IFP > 30 pMpc, finding that the luminosity dependency of IFP disappears. This reassures us that this trend is due to QSOs with low values of IFP, mainly at low luminosities, which could be possibly affected by associated absorbers. However, we cannot exclude that this effect is due to some biases caused by color selection effects which depend on the observed magnitudes. A further investigation is needed in the future to completely explain the observed correlation.**

In order to be sure that the examined QSO sample did not present too much different features from that of Prochaska et al. (2009) or Worseck et al. (2014), the stacking technique was also implemented to infer the evolution of the mean free path in redshift. The mean free path was computed from the rest-frame stacked spectra in different redshift bins of $\Delta z = 0.2$ from $z = 3.6$ to $z = 4.6$, modeling the observed flux in the region blue-ward of the Lyman limit as $f = f_{912} \exp(-\tau_{\text{eff,LL}})$, taking into account the spectral slope α_ν and the softening of the QSO spectrum at $\lambda_{rest} \lesssim 1000$ by Cristiani et al. (2016). In particular, the stacked spectra were averaged without weighting, in order not to introduce biases caused by sources without strong LLS absorption, which have higher S/N ratio (Prochaska et al. 2009). The MFP from the stack spectra were obtained by calculating the λ_{LLS} at which $\tau_{\text{eff,LL}} = 1$, as described in section 3.4, and converting this wavelength into a redshift z_{LLS} by assuming a rest-frame $\lambda_{rest} = 912$ Å.

As discussed in Section 3.4 (and in the forward in Section 6.1.1), the MFP estimation for QSOs at $3.6 < z < 3.8$ could be biased low due to the limited coverage of the observed SDSS spectra ($\lambda_{obs} \geq 3560$ Å). We have thus excluded the redshift bin $3.6 < z < 3.8$ from the fit in Fig. 10 and obtained steeper evolution of the MFP with redshift (Table 1, bottom part). As shown by our results, the redshift evolution is still milder than the one by Worseck et al. (2014).

If we take out from our sample all the QSOs with IFP ≤ 10 pMpc, assuming that they are affected by associated absorptions ("option 2"), the trends highlighted above are further migrating towards a milder evolution of MFP with redshift. Fig. 13 shows the updated mean values of IFP, divided in different bins of absolute magnitude, versus redshift. It is immediately evident that the mean values of the IFP are significantly higher than the Worseck et al. (2014) best fit (MFP ~ 50 pMpc at $z = 4$ vs MFP = 37 ± 2 pMpc by Worseck et al. (2014)) and the redshift evolution is much milder. Table 2 summarizes the best-fit values of our mean free path with a power law relation. At all luminosities the slope is $\eta \sim -2.6 \div -4.3$, which is significantly milder than the Worseck et al. (2014) one ($\eta = -5.4 \pm 0.4$).

We have also checked the effect of eliminating BAL QSOs from our sample. First of all, it is worth noting that the analysis carried out by Worseck et al. (2014) could include sources with possible associated absorbers (AAs), but excludes all the BAL QSOs. We have found that both the normalization and the slope of the MFP(z) relation are not affected by the presence of
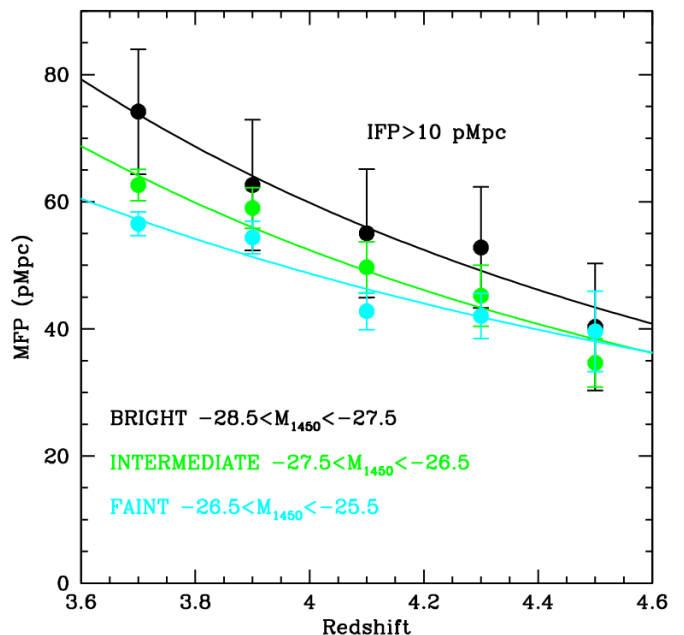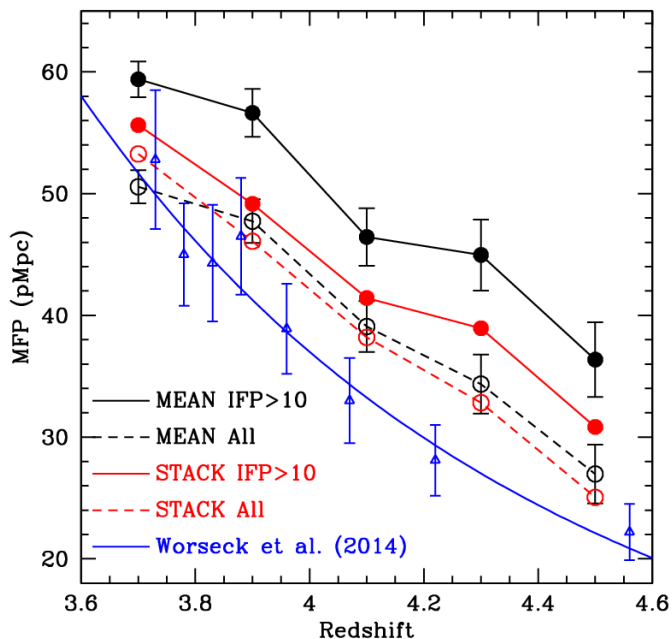

Fig. 13. The mean values of the IFP for SDSS QSOs at $3.6 \leq z \leq 4.6$, $I \leq 20.0$, and IFP > 10 pMpc ("option 2"), divided in redshift for three bins of absolute magnitude M_{1450}. The vertical bars are the standard error of the mean in each bin of absolute magnitude. The continuous lines show the best power-law fit to the observed mean points for each M_{1450} bin.

Table 2. Redshift evolution of the mean free path for QSOs at $3.6 \leq z \leq 4.6$ with IFP > 10 pMpc ("option 2"). We have excluded QSOs with IFP ≤ 10 pMpc (possibly affected by associated absorbers) from the fit of the MFP(z) relation. In the bottom part, we have excluded the $3.6 < z < 3.8$ redshift bin from the fit of the MFP(z) relation, since it could be affected by a possible underestimation of the IFP due to the limited wavelength coverage of the SDSS spectra in the bluer wavelengths.

η	MFP($z = 4$) pMpc	Magnitude bin M_{1450}
	All redshift bins	
-3.4 ± 0.4	59.8 ± 1.4	$-28.5 \leq M_{1450} \leq -27.5$
-3.3 ± 0.5	52.4 ± 1.4	$-27.5 \leq M_{1450} \leq -26.5$
-2.6 ± 0.6	48.7 ± 1.4	$-26.5 \leq M_{1450} \leq -25.5$
	Excluding 3.6 - 3.8 redshift bin	
-3.2 ± 0.8	59.4 ± 2.5	$-28.5 \leq M_{1450} \leq -27.5$
-4.3 ± 0.5	54.4 ± 1.2	$-27.5 \leq M_{1450} \leq -26.5$
-3.3 ± 1.0	49.7 ± 1.8	$-26.5 \leq M_{1450} \leq -25.5$

BAL QSOs, and the best fit values of the sample without BAL QSOs (and without AAs) are very similar to the ones in Table 2. This effect is not surprising, since both the LyC escape fraction and the individual free path distributions do not depend on the BAL QSO fraction. This indicates implicitly that there is no apparent relation from the BAL phaenomenon and the associated absorbers.

Fig. 14 summarizes the main results of this section. If we consider all the QSOs at $3.6 \leq z \leq 4.6$ and $I \leq 20.0$ ("option 1"), the mean value of the PDF(IFP) and the stack give similar results (as we discuss in Appendix B), while if we exclude QSOs with IFP ≤ 10 pMpc, assuming they are affected by associated absorbers ("option 2"), we obtain a huge difference between mean and stack and a milder redshift evolution w.r.t.

Fig. 14. The redshift dependence of the MFP for SDSS QSOs at $3.6 \leq z \leq 4.6$ and $I \leq 20.0$. The filled black points (connected by a continuous black line) summarize the mean value of the free paths excluding objects with IFP ≤ 10 pMpc, which are possibly affected by associated absorbers ("option 2"). The open black circles (connected by a dashed black line) show the MFP for all the QSOs of our sample ("option 1"). The filled red points (connected by a continuous red line) instead are derived from the stack of QSOs with IFP > 10 pMpc, and the open red circles (connected by a dashed red line) show the MFP obtained by the stack of all the QSOs. The dotted blue curve is the evolution of the MFP with redshift found by Worseck et al. (2014), while the blue open triangles indicate their mean values in different redshift bins.

Worseck et al. (2014). Table 3 summarizes the MFP obtained through stack and mean values of the PDF(IFP), compared with the results of Worseck et al. (2014).

The differences between our best fit and the Worseck et al. (2014) results could be due to two main reasons: 1-we exclude from our sample the QSOs which are suspected to be affected by associated absorbers, with IFP ≤ 10 pMpc ("option 2"); 2-we have used the mean values of the PDF(IFP), instead of the stack as done by Prochaska et al. (2009) and Worseck et al. (2014). As discussed in the Appendix B, an underestimation of the MFP from stacking compared to the mean value of the PDF is expected, in particular for skewed distributions as we are finding at $z \gtrsim 4$. Following these arguments, the probability distribution function of individual free path of UV ionizing photons can give more information on the IGM than the stacking technique. For this reason, in the following we will use the PDF(IFP) in order to compute the contribution of QSOs/AGNs to the HI ionizing background.

From Fig. 13 it is interesting also to note that the differences between the bright and faint QSO samples is reducing at $z > 4.2$ and it seems almost negligible at $z \sim 4.5$. Moreover, the mean value of the IFP distributions is progressively diverging from the Worseck et al. (2014) estimates from $z = 3.6$ to $z = 4.6$ for "option 2". This could have deep implications for the MFP estimation at the epoch of reionization. For example, from Worseck et al. (2014) relation a mean free path of ~ 6 pMpc is expected at $z = 6$, while a MFP three times larger (MFP ~ 18 pMpc)

is awaited extrapolating our relation in Table 3 and Fig.14. If the outlined trend shown in Fig.13 for "option 2" is kept also at higher redshifts, this could have important implications for the computation of the photo-ionization rate at the EoR, as we will discuss in the next sub-section.

5.3. The corrected ionizing background at z=4 and implications on Reionization

The contribution of bright QSOs, faint AGNs and/or star-forming galaxies to the HI ionizing background can be quantified by computing the photo-ionization rate of neutral hydrogen atoms, Γ_{HI}. This parameter primarily depends on the overall emissivity of the sources and on their mean free path, i.e. $\Gamma_{HI} \propto \text{MFP} \cdot f_{esc} \cdot \epsilon_{912}$, where f_{esc} is the escape fraction of Lyman continuum photons and ϵ_{912} is the overall intrinsic contribution of the ionizing sources emitted at 912 Å rest-frame. The latter depends on the luminosity function of high-z AGNs or SFGs and on the spectral slope of the sources close to 912 Å rest-frame. Since it is not the aim of this paper to revise the estimate of the emissivity of the different populations of galaxies and AGNs, we will only compute here the scaling factor due to our revised estimates of the MFP and f_{esc} at $z \sim 4$, $\Gamma_{HI}^{new}/\Gamma_{HI}^{old}$.

If the main sources of HI ionizing photons are star-forming galaxies at high redshift, the scaling factor for Γ_{HI} is simply the ratio $\text{MFP}_{mean}/\text{MFP}_{W14}$, from Table 4. It varies from 0.98 to 1.27, if all the QSOs are used for an estimate of the mean free path ("option 1"). The correction is significantly higher if the mean free path is derived only by QSOs with IFP > 10 pMpc ("option 2"). In the latter case, the ratio $\text{MFP}_{mean}/\text{MFP}_{W14}$ increases from ~ 1.1 at $z = 3.7$ to 1.6 at $z = 4.5$.

In the opposite case that QSOs and AGNs are the main producers of the ionizing background at $z \gtrsim 4$, the ratio $\Gamma_{HI}^{new}/\Gamma_{HI}^{old}$ does not simply scale as $\text{MFP}_{mean}/\text{MFP}_{W14}$. We should take into account two additional correction factors, i.e. 1-the possible correlation between individual free path and escape fraction (see Fig.7) and 2-the fraction of QSOs with IFP greater than 10 pMpc for "option 2". The first factor ($CORR1$ in Table 4) has been computed as $< \text{IFP} * f_{esc} > / (< \text{IFP} > * < f_{esc} >)$ and is always greater than 1, given the observed correlation between these two parameters (see Fig.7). The fraction $CORR2$ of QSOs with IFP > 10 pMpc, instead, is of the order of 0.7-0.8 for "option 2", while $CORR2$ is always 1.0 for "option 1", by definition.

If we consider all the QSOs ("option 1"), then the new estimate of MFP_{mean} is not significantly larger than previous estimates, but this is balanced by a strong correction for the correlation between IFP and escape fraction ($CORR1 \sim 1.2 - 1.3$). In this case the global correction is of the order of $\sim 1.2 - 1.7$ (upper part of Table 4).

If we instead exclude from our calculations QSOs with IFP ≤ 10 pMpc ("option 2"), the higher value of MFP is compensated by the fact that only 70-80% of the lines of sight are contributing to the photo-ionization rate. In this case (bottom part of Table 4) the total correction to Γ_{HI} is 1.2 in the redshift interval $z = 3.6 - 3.8$, rising to 1.4-1.7 at higher redshift. In addition, this ratio increases from $z = 3.7$ to $z = 4.5$. Interestingly, the global correction to the ionizing background $\Gamma_{HI}^{new}/\Gamma_{HI}^{old}$ does not depend whether QSOs with IFP ≤ 10 pMpc are included ("option 1") or excluded ("option 2") in our calculations, but it is practically consistent between the two cases.

It is worth noting that the values of $\text{MFP}_{mean}/\text{MFP}_{W14}$ inferred in this work are only lower limits. In fact, the averaged values of the mean free path computed from the IFP distribu-

Table 3. Summary of the MFP vs redshift. In the first column are the redshift bins. MFP$_{W14}$ indicates the MFP from the best-fit formula by Worseck et al. (2014). MFP$_{stack}^{all}$ and MFP$_{mean}^{all}$ are the values obtained by the stack technique and the mean of the distribution functions, respectively, for the whole sample ("option 1"). MFP$_{stack}^{noAAs}$ and MFP$_{mean}^{noAAs}$ are the same quantities excluding QSOs with IFP $\leq$ 10 pMpc ("option 2"). The last two columns are the ratio between the MFP values obtained by the distribution functions for "option 1" and "option 2", respectively, and MFP$_{W14}$.

Redshift bin	MFP$_{W14}$ pMpc	MFP$_{stack}^{all}$ pMpc	MFP$_{mean}^{all}$ pMpc	MFP$_{stack}^{noAAs}$ pMpc	MFP$_{mean}^{noAAs}$ pMpc	$\frac{MFP_{op1}}{MFP_{W14}}$	$\frac{MFP_{op2}}{MFP_{W14}}$
3.60-3.80	51.68	53.26	50.56	55.63	59.39	0.978	1.149
3.80-4.00	41.27	46.08	47.74	49.17	56.64	1.157	1.373
4.00-4.20	33.25	38.20	39.09	41.43	46.44	1.176	1.397
4.20-4.40	27.01	32.83	34.36	38.95	44.96	1.272	1.665
4.40-4.60	22.11	25.06	26.97	30.83	36.37	1.220	1.645

Table 4. Total corrections to the HI photo-ionizing background. The redshift bins are indicated in the first column. MFP$_{mean}$/MFP$_{W14}$ is the ratio between our value MFP$_{mean}$ and MFP$_{W14}$, which are taken from Table 3. N_{QSO} is the number of sources in each redshift bin with IFP above a given threshold. $CORR1$ is the correction term which takes into account the correlation between escape fraction and IFP, as observed in Fig.7, and it is computed as $<IFP * fesc> / (<IFP> * <fesc>)$. $CORR2$ is the fraction between the number of QSOs with IFP above a given threshold and the total number of QSOs in each redshift bin. $\Gamma_{HI}^{new} / \Gamma_{HI}^{old}$ is the product of MFP$_{new}$/MFP$_{W14}$ with $CORR1$ and $CORR2$.

Redshift bin	$\frac{MFP_{mean}}{MFP_{W14}}$	N_{QSO}	$CORR1$	$CORR2$	$\Gamma_{HI}^{new} / \Gamma_{HI}^{old}$
		Option 1: all QSOs			
3.6-3.8	0.978	1099	1.283	1.000	1.255
3.8-4.0	1.157	668	1.258	1.000	1.455
4.0-4.2	1.176	382	1.314	1.000	1.545
4.2-4.4	1.272	219	1.370	1.000	1.742
4.4-4.6	1.220	140	1.402	1.000	1.710
		Option 2: only QSOs with IFP > 10 pMpc			
3.6-3.8	1.149	915	1.309	0.833	1.252
3.8-4.0	1.373	550	1.283	0.823	1.450
4.0-4.2	1.397	312	1.350	0.817	1.540
4.2-4.4	1.665	159	1.428	0.726	1.726
4.4-4.6	1.645	96	1.492	0.686	1.683

tion functions may be even greater if all the possible caveats in the estimation of this parameter are taken into account (i.e. observational limitations and selection effects of some QSOs, see Section 6.1).

In summary, our results indicate a milder evolution of the MFP with redshift, giving a revised estimates of the photo-ionization rate by a factor of 1.4-1.7 upward at $z \sim 4.0 - 4.5$. If the trend obtained with these QSOs is extrapolated at $z \sim 6$, then the new estimates of the MFP, and consequently of the ionizing background, could be a factor of 2-3 times higher than the recent estimates, both for star-forming galaxies and QSOs/AGNs (e.g. Finkelstein et al. (2019); Giallongo et al. (2019)). In the future it will be important to check whether this trend is still present at $z > 5$.

# 6. Discussions

6.1. Possible systematic effects

This work is based on a sample of $z \sim 4$ QSOs selected by the SDSS thanks to their optical photometry (Paris et al. 2018). The specific selection criteria adopted by SDSS may affect our results, especially for what concerns the escape fraction and mean free path estimations. We take into account here the effect of the

limited extension of SDSS spectra in the UV and the color selections of $z \sim 4$ QSOs. Both these issues could prevent a correct and robust estimation of the total contribution of AGNs to the HI ionizing background at $z \sim 4$. We do not attempt here to correct for these two effects, but it is easy to conclude, from what we are going to show in the following, that the net effect of these two issues leads to an underestimate of both the escape fraction and MFP. Dedicated observations and well-defined surveys are needed at the end to properly correct for these problems.

6.1.1. The limited coverage of SDSS spectra in the UV

The SDSS spectra released in the DR14 distribution are limited to the wavelength range $3600 \lesssim \lambda_{obs} \lesssim 10000$ Å. The cut at lower wavelengths in the UV can affect our estimation of the free path of QSOs at $z \sim 4$, as shown in Fig. B.1. In this case, this QSO (SDSS J113654.6+485322.3 at $z_{spec} = 3.61964$) has a long path along the line of sight clear from damped Lyman-α or Lyman limit systems. Its ionizing radiation is thus able to travel tens of proper Mpc without being attenuated by a factor $1/e$ (defining the IFP) and can reach the observed lower limit of the SDSS spectrum at $\lambda_{obs} \sim 3550$ Å. It is not known, with the present data, whether the ionizing radiation of this QSO is able to reach bluer wavelengths, and for this reason an individual free path of 145.94 pMpc is assigned. This is a lower limit, simply imposed by an instrumental limitation; the real value of the IFP for this object could be larger than the value assigned here. **In our sample, for 12.8% of the QSOs the provided IFP is a lower limit due to this effect. In the redshift bin** $3.6 < z < 3.8$ **this fraction is 15.4% and it is progressively decreasing to 10% in the highest redshift bins considered.** In the future it will be important to spectroscopically follow up these kind of objects to definitively measure their IFP and to ensure that the SDSS spectra are not affected by any systematics.

This issue causes the artificial lack of QSOs with long IFP at lower redshifts, especially close to $z \sim 3.6$, as shown in Fig. 15. Many sources with high free paths could present a spectrum which extends far beyond the lower instrumental limit of the observed wavelengths; this results in an altered estimation of the IFP distribution. Since at $z \sim 4$ we are finding QSOs with IFP ~ 200 pMpc or larger, we do expect that these sources will be present also at lower redshifts. Their absence however is due only to instrumental limitation, as shown by Fig. 15. The blue line in this plot represents the maximum values of the IFP that can be obtained considering a minimum observable wavelength of about 3550 Å at different redshifts. These limits exactly follow the distributions of the sources with the highest IFP, confirming that this trend is not a physical effect, but an instrumental feature, as explained before. To correct for this systematic effect the
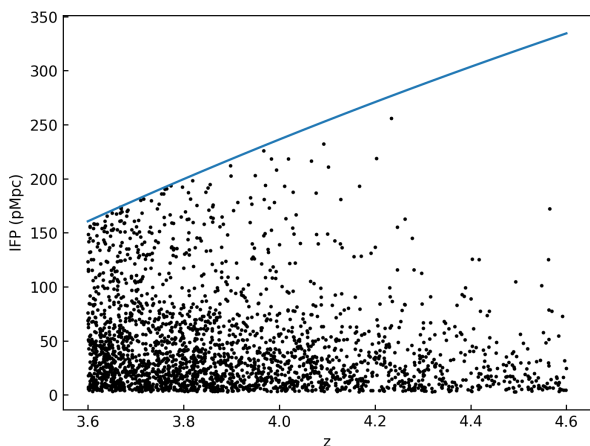
Fig. 15. Individual free path (black points) vs spectroscopic redshift for SDSS QSOs at $3.6 \leq z \leq 4.6$. The blue line represents the maximum value of IFP which can be obtained at each redshift due to the observational limits of the SDSS spectra, i.e. $\lambda_{obs} \gtrsim 3550$ Å.

sources close to the instrumental limit shown in Fig. 15, deep spectra with spectrographs sensitive to UV bluer wavelengths are required, i.e. MODS1 at LBT or FORS1 at VLT, as shown by Grazian et al. (2018).

In conclusion, this particular issue can artificially reduce the average value and the scatter of the free path obtained through the probability distribution function of the QSO sample, especially at $z \leq 3.9$. This effect is particularly noticeable in Fig. 12 and 13. For example, if we compute the MFP in the redshift interval $4.0 \leq z \leq 4.1$, we obtain MFP = 40.73 pMpc, while, if we exclude sources with IFP $\geq$ 150 pMpc (thus simulating the bias observed at $z \sim 3.6$), the average value is artificially lower, MFP = 34.61 pMpc. Thus, the MFP adopted in the present paper at $z \leq 3.9$ could be higher, of the order of $\sim$ 18%.

This underestimation of the mean free path at $z \leq 3.7$ can also artificially flatten the redshift evolution of this parameter shown in Fig.14. We have excluded the first redshift bin and compute the best fit for the sample of QSOs with IFP > 10 pMpc ("option 2"), and find $\eta \sim -4.3 \div -3.2$, with slightly larger uncertainties ($\sigma_\eta \sim 0.8$). In the future, deep spectra extending in the UV at $\lambda_{obs} \sim 3300$ Å for these QSOs would provide a more robust estimation of the redshift evolution of the MFP.

6.1.2. Color Selection Effects

The selection criteria adopted by the SDSS team to select the QSOs analyzed in this paper are based on optical colors (Paris et al. 2018). In particular, the selection of QSOs at $3.6 \leq z \leq 4.2$ is mainly based on the color criterion $u - g > 1.5$ (Richards et al. 2002), which is especially useful because it involves the Gunn filters u and g, centered at the average wavelengths 3551 and 4686 Å, respectively. The $u - g$ color takes advantage of the expected flux decrement of the source in the region blueward of the Lyman limit. Problems arise when the observed QSO presents a large individual free path. This effect causes a bluer $u - g$ color, which eventually could be lower than the SDSS threshold used to select QSOs at $z > 3.6$. Fig. 16 shows the effect discussed here. In this plot, our QSO sample has been split into two parts using a threshold in IFP of 150 pMpc, in the cases of sources with higher and lower redshift w.r.t. the median

one of the sample ($z \sim 3.90$). The $u - g$ distribution with larger (smaller) free paths with respect to the threshold is represented by the green (orange hatched) histogram, and its mean value is marked by the dot-dashed (solid) black line. As can be seen, at larger values of the IFP, the mean value of the distribution moves to bluer $u - g$ colors (green histograms), approaching the 1.5 threshold adopted by SDSS to select the observed QSOs. In particular, this effect is more pronounced at lower redshift (bottom panel) where a possible lack of sources with high IFP would change the evolution of this parameter with redshift. For these reasons, QSOs with large IFP could be lost in the standard selection procedure adopted by SDSS, perhaps artificially decreasing the average value and scatter of the free path distribution and thus the overall contribution of these sources to the UV ionizing photon production. This effect is particularly severe at $z \lesssim 3.5$, as been pointed out by Prochaska et al. (2009) and Cristiani et al. (2016).

It is useful to note here that the few QSOs at $3.6 \leq z \leq 4.6$, with $u - g \leq 1.5$ shown in Fig. 16 (123 out of 2508 sources in the all redshift range), have been observed by SDSS thanks to other selection criteria, complementary to the optical color selection of SDSS, as described by Dawson et al. (2013); Paris et al. (2018). It is thus not possible to estimate the systematic effects due to this color selection criterion with a simple analytic calculation. Just as an example, taking into account the 16 faint AGNs found at $3.6 \leq z \leq 4.2$ in the COSMOS field by Boutsia et al. (2018), only 5 (approximately 30%) would be selected adopting the color selection of SDSS, with an incompleteness of $\sim$70%. Interestingly, the missed AGNs are selected mainly by X-ray emission (Civano et al. 2016; Marchesi et al. 2016) and show large values of LyC escape fraction (Grazian et al. 2018) and individual free path. A dedicated survey, unbiased towards high IFP sources, is required in the future to solve this issue, which could be particularly severe, both for the proper estimate of the average value of the free path (and its scatter) and for a correct quantification of the number density of QSOs at high redshifts. This survey, however, will have the severe issue of strong contamination by nearby galaxies and stars, which are affecting the samples selected by adopting bluer $u - g$ colors.

6.2. The population of $z \sim 4$ QSOs with low escape fractions and short free paths

In the previous sections we have chosen a threshold above 10 pMpc for the individual free path in order to study the properties of the IGM. We suspect indeed that the QSOs with a free path lower than this limit could be affected by associated (intrinsic) absorbers. At this aim, it is worth noting here that we do not claim that the associated absorbers are acting on scales of 10 pMpc. Our hypothesis is that associated absorbers residing few kpc away from the QSO can have large peculiar velocities (outflows up to $\sim 10^4$ km s^{-1}) and thus they can mimic a cosmological distance of $\lesssim$10 pMpc from the QSO itself.

With the aim of verifying this guess, we have cross-correlated our sample with the XQ-100 sample (Lopez et al. 2016), which is a legacy survey of one hundred quasars at $3.5 < z < 4.5$ observed with VLT/X-shooter for a deep exposure time. We find 44 objects in common, and we then refer to Perrotta et al. (2016) in order to study the properties of their absorbers. The latter find that there is a statistical significant (up to 8 sigma) excess of associated NV absorbers at a velocity of $|v| \leq 10^4$ km s^{-1}, corresponding to $\sim$ 20 pMpc. This is slightly different with respect to our adopted threshold of 10 pMpc, but it can be comparable to it, considering the redshift uncertainties
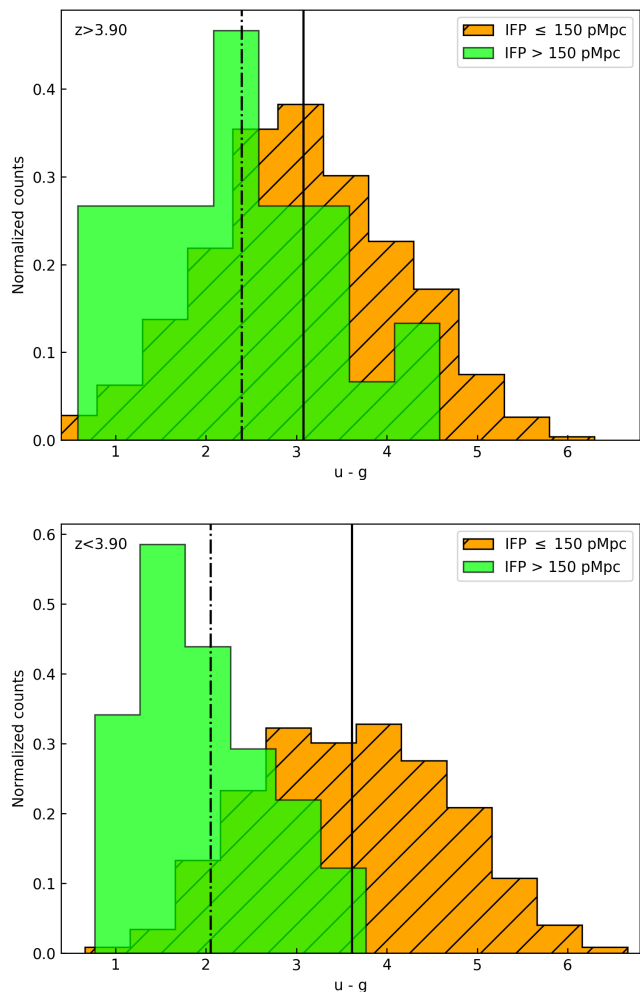
Fig. 16. Distribution functions of the $u - g$ color for QSOs with high (green histogram, IFP $\geq$ 150 pMpc) and low (orange hatched histogram, IFP < 150 pMpc) individual free paths. The solid and dot-dashed black lines refer to the average values of the orange and green distributions, respectively. The top (bottom) panel shows the $u - g$ distributions for QSOs at higher (lower) redshift w.r.t. the median one at $z \sim 3.90$.

of $\sim$ 0.005 for the SDSS sample and the intrinsic uncertainties on the IFP determination of 2-3 pMpc, due to our procedure.

Out of the 44 QSOs in common between our sample and the XQ-100 one, we check in detail the absorbers of the 6 QSOs with free path less than 10 pMpc. We find that, for all six, the C_{IV} absorbers studied by Perrotta et al. (2016) lie within 1-2 pMpc from our estimates, which indicates these absorbers are the responsible of the short free path measured by our methods, and of the low value of the escape fraction, as we have discussed above. More precisely, 2 out of 6 QSOs have a NV system, which according to Perrotta et al. (2016) is a robust signature of associated absorbers. Indeed, our recovered NV fraction (33%) is similar to the ones found by Perrotta et al. (2016) and Perrotta et al. (2018), i.e. 33% and 38%, respectively. The other four systems have no individual detection of NV in Perrotta et al. (2016), however a more recent work by Perrotta et al. (2018) indicates that C_{IV} absorptions with $N(C_{IV}) \geq 10^{14} \text{cm}^{-2}$ within 5000 km s^{-1} of the emission redshift of the QSO (mimicking a distance of $\sim$ 10 pMpc) are detected in NV at 15 σ through stacking, supporting our guess that probably all these systems are bona-fide associated, given their large column densities and

high excitations. High ionization absorption lines indeed require close proximity to the QSOs based on detailed photo-ionization constraints (Chen et al. 2018).

7. Summary and Conclusions

The aim of this paper is to measure the LyC escape fraction and the free path of HI ionizing photons for a complete sample of the $z \sim 4$ QSO population, as a follow-up study of the work by Cristiani et al. (2016). The f_{esc} and MFP are fundamental parameters involved in the characterization of the IGM and, in particular, in the estimation of the HI photo-ionization rate Γ_{HI} at high redshifts. At this aim, we select 2508 QSOs at $3.6 \leq z \leq 4.6$ and in the magnitude range $17.0 \leq I \leq 20.0$ ($-29.0 \lesssim M_{1450} \lesssim -26.0$).

In order to quantify the amount of UV photons able to ionize the surrounding IGM, the LyC escape fraction of each QSO was computed as the ratio between the average flux blue-ward ($892 \leq \lambda_{rest} \leq 905$ Å) and red-ward ($915 \leq \lambda_{rest} \leq 960$ Å) of the Lyman limit. The PDF(f_{esc}) resulted in a multi-modal distribution with a mean value $f_{esc} = 0.49 \pm 0.36$. In particular, the narrow peak at $f_{esc} \leq 0.2$ could depend on the wavelength coverage used to compute the average flux in the ionizing region, blue-ward of the Lyman limit. In fact, a shorter spectral window, closer to 912 Å rest-frame, enhances the number of sources with high escape fraction, reducing the height of the peak at low f_{esc} (Fig. 6). This, in turn, increases the average value of the PDF(f_{esc}), making the result obtained in this paper just a lower limit to the real value of the f_{esc}. Indeed, the mean f_{esc} we find ($\sim$ 0.5) is consistent with the average flux decrement between 930 and 900 Å rest-frame due to intervening IGM absorption derived by Inoue et al. (2014) at $z = 4$ ($\sim$ 0.6, see their Fig. 4), suggesting that the escape fraction could likely reach 100% in bright QSOs, if we correct for this effect. This last consideration is then strengthened by the connection between f_{esc} and the individual free path of HI ionizing photons we find in this paper. Sources with IFP $\leq$ 10 pMpc also show a low LyC escape fraction (Fig. 7); as reported in Section 5, this is again an effect of the spectral window chosen to compute f_{esc}. Furthermore, QSOs with IFP $\leq$ 10 pMpc (possibly affected by associated absorbers) should not be ideal tracers of the properties of the IGM. Finally, a redshift and magnitude evolution of f_{esc} was excluded. Therefore, the ionizing properties of faint AGNs and bright QSOs seem to be the same at $z \sim 4$, in agreement with Grazian et al. (2018).

Regarding the estimation of the mean free path of HI ionizing photons from its distribution function, this is a method which could represent an improvement with respect to the analysis performed so far in the literature. Indeed, the classical approach to the calculation of the mean free path relies on assumptions on the frequency distribution of IGM absorbers or on the direct estimation of this parameter from the stacked spectra of QSOs in different redshift bins. On one side, the first approach is uncertain because of several issues on the derivation of the exact HI frequency distribution; on the other side, the stack technique seems to depend on the fraction of QSOs with low individual free paths. For this reason, the statistical analysis of the IFP distribution has been considered a more informative method to study the ionization status of the IGM.

Summarizing, we have found that: 1-the IFP distribution function carries more information than the MFP from stacking; 2-the MFP at $z \sim 4$ are larger than the one found by Worseck et al. (2014) by a factor of 1.1-1.7, probably due to their sam-

pling fainter luminosities at higher redshifts and not cleaning their sample for possible associated (intrinsic) absorbers; 3-the redshift evolution of the MFP we derive is milder than the one obtained by Worseck et al. (2014); 4-if the trends found at $z \sim 4$ and $M_{1450} \sim -27$ are confirmed also at higher redshifts and at fainter luminosities, this could have strong implications for the role of bright QSOs and faint AGNs on the hydrogen reionization process.

As we did for the LyC escape fraction, we check the dependence of the free path from luminosity and redshift. For the first case, we divide the sample in redshift bins of $\Delta z = 0.2$, each of which has been in turn split in magnitude bins $\Delta M_{1450} = 1.0$. The IFP of bright QSOs turns out to be larger than the one of faint QSOs, at least up to $z = 4$. This trend becomes milder going at higher redshifts and seems not to be affected by systematic effects.

As previously said about the free path, we used the same redshift and absolute magnitude bins to test its redshift evolution by computing the MFPs from the PDF in each bin; then, we compared our results with the ones of Worseck et al. (2014) w.r.t. which we found a milder trend of MFP versus redshift. This is mainly due to the technique involved in the calculation of the MFP by Worseck et al. (2014), i.e. the stack method, which could give an underestimation of the real value of the MFP at different redshifts (see Appendix B). Excluding QSOs with IFP ≤ 10 pMpc, assuming that these sources are affected by associated absorbers, makes the difference between the MFPs from the PDFs and the ones obtained by the stack even larger (MFP = 49 - 59 pMpc at $z = 4$, from our work, w.r.t. MFP = 37 ± 2 pMpc, by Worseck et al. (2014)), with milder evolution with redshift ($\eta \sim -2.6 \div -4.3$ versus the best fit of Worseck et al. (2014), i.e. $\eta = -5.4$). Extrapolating these results at higher redshifts could have important implications about the reionization process.

In this regard, we have computed the scaling factor $\Gamma_{HI}^{new}/\Gamma_{HI}^{old}$ in order to correct previous estimates of the photo-ionization rate according to our new measurements of MFPs from the PDFs. To do that, we consider two extreme cases: 1-all the QSOs with IFP ≤ 10 pMpc were affected by intervening absorbers ("option 1"); 2-that these latter sources were all associated (intrinsic) systems which should be rejected for a correct estimates of the mean free path ("option 2"). In both cases, for an ionizing background mainly produced by QSOs and AGNs, the correction factor goes from 1.2 at z=3.7 to 1.7 at z=4.5, respectively, with a possible monotonic trend with redshift. It is worth noting that recent results of Perrotta et al. (2016, 2018) indicate that at least 30% of the CIV absorbers at $|v| \leq 10^4 km s^{-1}$ (i.e. at $\leq 10 - 20$ pMpc if we translate the velocity offset into a cosmological distance) from the QSOs could be affected by associated (intrinsic) systems.

Finally, it is worth noting that our estimation of the PDF of the free path at $z \geq 3.5$ is of paramount importance since it gives a reference benchmark for the theoretical models which are investigating the reionization epoch with state-of-the-art cosmological hydrodynamical simulations with complex radiative transfer calculations. It will be interesting to check whether these models are able to reproduce the mean value, the scatter, and the redshift evolution of the IFP distribution functions outlined in this paper. Detailed comparison with state-of-the-art simulations, as the ones carried out by Inoue et al. (2014) will be the subject of a future paper.

Although these ones are important clues for studying the evolution of the MFP with redshift, it is necessary to go deeper in magnitude, by observing fainter AGNs at $z \geq 4$ in order to increase the statistics on these sources and thus to better investi-

gate their possible luminosity evolution. This will be done by a follow-up survey, extending the seminal work done by Grazian et al. (2018). In addition, extending in redshift at $z \sim 5 - 6$ the analyzed sample of QSOs is of paramount importance to confirm the trends we have found at $z \leq 4.6$.

References

Akiyama, M., He, W., Ikeda, H., et al. 2018, PASJ, 70S, 34
Becker, G. D., Bolton, J. S., Madau, P., et al. 2015, MNRAS, 447, 3402
Bolton, A. S., Schlegel, D. J., Aubourg, E., et al. 2012, AJ, 144, 144
Boutsia, K., Grazian, A., Giallongo, E. et al. 2018, ApJ, 869, 20
Calverley, A. P., Becker, G. D., Haehnelt, M. G., & Bolton, J. S., 2011, MNRAS, 412, 2543
Cardelli, J. A., Clayton, G. C., Mathis, J. S., 1989, ApJ, 345, 245
Chen, C., Hamann, F., Simon, L., et al. 2018, MNRAS, 481, 3865
Civano, F., Marchesi, S., Comastri, A., et al. 2016, ApJ, 819, 62
Cowie, L. L., Barger, A. J., Trouille, L., 2009, ApJ, 692, 1476
Crighton, N. H. M., Prochaska, J. X., Murphy, M. T., et al. 2019, MNRAS, 482, 1456
Cristiani, S., Serrano, L. M., Fontanot, F., Vanzella, E., & Monaco, P. 2016, MNRAS, 462, 2478
D'Aloisio, A., McQuinn, M., Davies, F. B., & Furlanetto, S. R., 2018, MNRAS, 473, 560
Davies, F. B., Hennawi, J. F., Eilers, A.-C., Lukić, Z., 2018, ApJ, 855, 106
Dawson, K. S., Schlegel, D. J., Ahn, C. P., et al. 2013, AJ, 145, 10
Dayal, P. & Ferrara, A. 2018, Physics Reports, 780, 1
Fan, X., Strauss, M. A., Schneider, D. P., et al. 2001, AJ, 121, 54
Fan, X., Strauss, M. A., Becker, R. H., et al. 2006, AJ, 132, 117
Faucher-Giguère, C.-A., Lidz, A., Hernquist, L. & Zaldarriaga, M., 2008, ApJ, 688, 85
Faucher-Giguère, C.-A., 2019, arXiv:1903.08657
Finkelstein, S. L., D'Aloisio, A., Paardekooper, J.-P., et al. 2019, ApJ subm., arXiv:1902.02792
Fontanot, F., Cristiani, S. & Vanzella, E., 2012, MNRAS, 425, 1413
Fumagalli, M., Prochaska, J. X., Kasen, D., et al. 2011, MNRAS, 418, 1796
Giallongo, E., Menci, N., Fiore, F., et al. 2012, ApJ, 755, 124
Giallongo, E. Grazian, A., Fiore, F., et al. 2015, A&A, 578A, 83
Giallongo, E., Grazian, A., Fiore, F. et al. 2019, ApJ accepted, arXiv:1909.00702
Grazian, A., Boutsia, K., Giallongo, E., et al. 2018, A&A, 613A, 44
Haardt, F., & Madau, P. 2012, ApJ, 746, 125
Inoue, A. K. & Iwata, I., 2008, MNRAS, 387, 1681
Inoue, A. K., Shimizu, I., Iwata, I., Tanaka, M., 2014, MNRAS, 442, 1805
Kim, Y., Im, M., Jeon, Y. et al. 2019, ApJ, 870, 86
Kollmeier, J. A., Zasowski, G., Rix, H.-W., et al., 2017, arXiv:1711.03234
Kulkarni, G., Worseck, G., and Hennawi, J. F., 2019, MNRAS accepted, arXiv:1807.09774
Lopez, S., D'Odorico, V., Ellison, S. L., et al., 2016, A&A, 594, 91
Lusso, E., Worseck, G., Hennawi, J. F., et al. 2015, MNRAS, 449, 4204
Madau, P. & Haardt, F. 2015, ApJL, 813, 8
Matsuoka, Y., Oyabu, S., Tsuzuki, Y., et al. 2005, PASJ, 57, 563

Matsuoka, Y., Onoue, M., Kashikawa, N., et al. 2018, PASJ, 70S, 35
Marchesi, S., Civano, F., Elvis, M., et al. 2016, ApJ 817 34
McQuinn, M., 2016, ARA& A, 54, 31
Meiksin, A. & White, M. 2004, MNRAS, 350, 1107
Meiksin, A., 2009, Reviews of Modern Physics, 81, 1405
O'Meara, J. M., Prochaska, J. X., Worseck, G. et al. 2013, ApJ, 765, 137
Oke, J. B. & Korycansky, D. J. 1982, ApJ, 255, 11
Paris, I., Petitjean, P., Aubourg, E., et al. 2018, A&A, 613A, 51
Parsa, S., Dunlop, J. S. & McLure, R. J. 2018, MNRAS, 474, 2904
Perrotta, S., D'Odorico, V., Prochaska, J. X., et al. 2016, MNRAS, 462, 3285
Perrotta, S., D'Odorico, V., Hamman, F., et al. 2018, MNRAS, 481, 105
Planck Collaboration, 2018, arXiv:1807.06209
Prochaska, J. X., Worseck, G., & O'Meara, J. M. 2009, ApJL, 705, 113
Prochaska, J. X., O'Meara, J. M. & Worseck, G. 2010, ApJ, 718, 392
Prochaska, J. X., Hennawi, J. F., Lee, K.-G., et al. 2013, ApJ, 776, 136
Prochaska, J. X., O'Meara, J. M., Fumagalli, M. et al. 2015, ApJS, 221, 2
Rahmati, A. & Schaye, J. 2018, MNRAS, 478, 5123
Richards, G. T., Fan, X., Newberg, H. J., et al. 2002, AJ, 123, 2945
Rudie, G. C., Steidel, C. C., Shapley, A. E. et al. 2013, ApJ, 769, 146
Sargent, W. L. W., Steidel, C. C., & Boksenberg, A. 1989, ApJS, 69, 703
Schindler, J.-T., Fan, X., McGreer, I. D. et al. 2018, ApJ, 863, 144
Schindler, J.-T., Fan, X., McGreer, I. D. et al. 2019, ApJ, 871, 258
Shull, J. M., Stevans, M. L., & Danforth, C. W. 2012, ApJ, 752, 162
Songaila, A., & Cowie, L. L. 2010, ApJ, 721, 1448
Stevans, M. L., Shull, J. M., Danforth, C. W., Tilton, E. M., 2014, ApJ, 794, 75
Telfer, R. C., Zheng, W., Kriss, G. A., et al., 2002, ApJ, 565, 773
Worseck, G., Prochaska, J. X., O'Meara, J. M., Becker, G. D., et al. 2014, MN-
    RAS, 445, 1745
Yang, J., Wang, F., Fan, X., et al. 2019, ApJ, 871, 199

Appendix A: Example of SDSS QSO spectra

In this appendix, we report another example (beyond the one previously shown in Fig. 2), which explains the procedure adopted to estimate the IFP for each SDSS QSO. In particular, Fig. B.1 is an example of the issue related to the estimation of the free path for several sources in our sample, i.e. the limited wavelength coverage of some SDSS spectra. Here the LyC emission extends at wavelengths bluer than the lower limit of the SDSS spectrum, i.e. $\lambda_{obs} = 3560$ Å, indicated by the dashed orange line, but does not intersect the blue line, which marks the $1/e$ level w.r.t. the flux at 912 Å rest-frame. In this case the estimated IFP of 145.94 pMpc is only a lower limit due to the instrumental limitations, and the true free path for this QSO could be even larger.

Appendix B: The mean free path from stack spectra and from the mean value of the probability distribution function

The results from the statistical analysis on the IFP distribution functions for the high-z QSOs highlight a possible disagreement between these latter values and those from the stacks. In principle, this could be caused by the shape of the individual free path distributions, which result to be highly asymmetric/skewed and wide at $z \geq 3.6$. To test this hypothesis, a comparison between the MFP from the stacked spectra and from the mean value of the distributions of sub-samples selected in narrow intervals of IFP in the redshift range $3.60 \leq z \leq 3.80$ was accomplished. Table B.1 summarizes the obtained results. As reported in the first three rows of this table, bins of 10 pMpc in IFP were considered, namely between 50-60, 90-100, and 140-150 pMpc (column 1); in each of these bins, a number N_{obj} of QSOs was found (column 2). Comparing the mean free path from the stacks and distributions (columns 3 and 4, respectively), it can be seen that the discrepancies are clearly reduced.

Furthermore, two distant peaks were also analyzed in order to study the effect of the asymmetry/skewness of the distribution on the stacked spectrum. The obtained values are reported in the last row of Table B.1 for two peaks at 15 and 105 pMpc, with 36 and 8 QSOs in each bin, respectively. In this case, the results from the stack and the distribution are quite different from each other, being the former value significantly lower than the latter one. These observations seem to lead to the conclusion that in a wide and really asymmetric distribution, the stack is dominated by sources with small IFP, resulting in a lower value w.r.t. the one obtained by the distribution function. Moreover, the stack technique seems to depend on the number of sources with low free path with respect to those with higher values of this parameter. Indeed, if the number of QSOs in the bin with low free paths is about or lower than twice of that in the highest bin, the stacked spectrum is dominated by these latter sources, resulting in the estimation of a large MFP. On the other hand, if the ratio between the two numbers is major than three (which is the case discussed here), sources with low free paths dominate the stack, which provides a result lower than the one computed from the mean of the associated distribution. This is a simple consequence of the definition of IFP, which is the distance where the emitted flux is reduced by a factor of $1/e \sim 1/3$, i.e. thus dominated by the bulk of the IFP distribution.

Finally, it is reasonable to believe that the disagreement between the two methods discussed above could increase going at higher redshift (i.e. $z > 4$) because of the likely more pronounced asymmetry of the IFP distribution functions. Certainly,
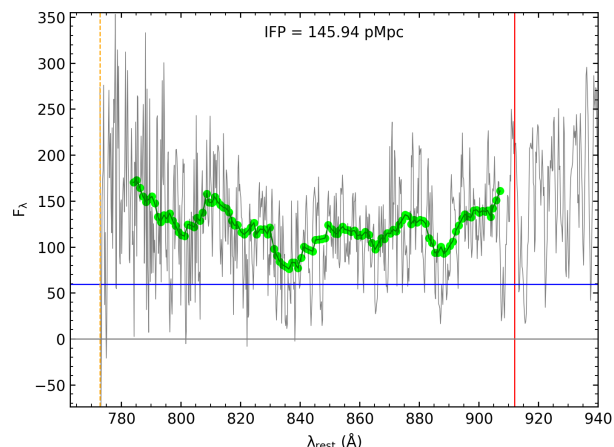
Fig. B.1. Free path estimation of the **QSO SDSS J113654.6+485322.3** at $z_{spec} = 3.61964$. Same colors legend as in Fig. 2.

Table B.1. Stack vs Probability Distribution Function from simulations. Comparison between the estimation of the MFP from the stacked spectra and the mean of the probability distribution functions. Narrow and distant bins were analyzed (first, second, and third rows) in order to study the effect of the asymmetry and of the width of the distributions on the stack technique. The studied bins, their number of sources, and the values obtained from the stacks and the distributions are reported from the first to the last column. The last row reports the results of the analysis of two distant bins at 15 and 105 pMpc with 36 and 8 QSOs, respectively, in order to study the effect of the asymmetry of the distribution on the stacked spectra.

MFP interval pMpc	N_{obj}	MFP$_{stack}$ pMpc	MFP$_{mean}$ pMpc
50-60	36	54.67	54.46
90-100	22	98.84	96.08
140-150	11	150.82	144.66
15/105	36/8	20.26	28.86

approaching the end of the reionization at $z \sim 6 - 7$, the discrepancy should be reduced, because of the decrement of the free paths for all the sources, resulting in a PDF peaked at low values.

The simulations carried out here are quite simplistic and it is possible that they are not able to fully reproduce the real IFP distributions. However, they are useful in order to understand the behaviour of the stacking procedure when applied to skewed distributions, as the ones shown in Fig. 6.