

Mémoire d’Habilitation à Diriger des Recherches

*From atomic to molecular gas along quasar lines of sight:
an absorbing point of view on galaxy evolution*

Pasquier Noterdaeme

CNRS - Institut d’Astrophysique de Paris

Habilitation à diriger des recherches soutenue le / 2019, en présence du jury composé de:

Alain Blanchard (Professeur UPS, IRAP)– *Rapporteur*

Jean-Gabriel Cuby (Astronome, LAM) – *Rapporteur*

Florence Durret (Astronome, IAP) – *Présidente*

Cécile Gry (Astronome, LAM) – *Rapportrice*

Christophe Yèche (Chercheur, CEA-IRFU) – *Examineur*

Remerciements

Au moment d'écrire les remerciements pour ce mémoire de HDR, je suis retourné chercher ma thèse de doctorat, quelque-part perdue sur une étagère, et relu, par curiosité, les remerciements qui s'y trouvaient. J'ai été presque surpris par le petit nombre de chercheurs mentionnés, c'est à dire mes directeurs de thèse, quelques collaborateurs ainsi que les membres du jury.

Dix ans plus tard, j'aurais besoin de plusieurs pages pour remercier toutes celles et ceux qui ont contribué à mon développement scientifique et participé aux recherches auxquelles je tient à coeur –et que je résume dans ce document. Il s'agit évidemment des chercheurs avec qui j'ai eu la chance de collaborer ces dernières années, mais aussi avec qui j'ai eu des discussions constructives, dans les couloirs de l'IAP, ou ailleurs, lors de missions ou de conférences, et avec qui j'ai parfois pu être en désaccord (ce qui est en fait très utile pour faire avancer les idées). Par peur d'oublier quelques noms, je préfère n'en donner aucun et donc les remercier collectivement, mais non pas moins chaleureusement (en principe, bibtex a bien fait le travail, et la plupart devraient se retrouver dans les références).

J'aimerais également remercier les référés (généralement anonymes) de mes différentes publications ainsi que les rapporteurs et membres du jury HDR d'avoir accepté de lire ce mémoire.

Enfin, je me dois de remercier les étudiants que j'ai pu encadrer ou co-encadrer, me donnant, peut-être parfois à leurs dépens, l'opportunité d'apprendre à transmettre des connaissances. J'espère que ce manuscrit y contribuera, au moins pour les suivants.

Depuis ma thèse de doctorat, ma famille s'est également agrandie, et mes plus tendres pensées vont à ma charmante femme Lorena et à nos trois adorables enfants: Luciano, l'aîné, rêveur inventif perdu dans les étoiles, Violeta, magique esthète mais aux pieds sur terre et enfin Facundo, le petit clown de la maison, qui nous a toujours fait rire même lorsqu'il passait lui-même de rudes épreuves (bien plus difficiles que la rédaction d'un mémoire de HDR). Quitte à parler d'épreuve, je souhaite également avoir une pensée pour mon père, chirurgien humanitaire, pour son amour de la vie, même si celle-ci l'a quitté brutalement l'an dernier.

Voilà, il ne me reste plus qu'à te remercier, toi, lecteur, de jeter un coup d'oeil aux pages qui suivent, vu que tu es déjà arrivé jusqu'à la fin des remerciements.

Contents

1	Introduction	7
1.1	Gas in nearby galaxies	7
1.2	Gas in distant galaxies	8
1.3	Quasars as cosmological probes of the gas	8
2	Atomic gas	11
2.1	Surveys	11
2.2	Distribution of the atomic gas in the Universe	12
2.2.1	Formalism and definitions	12
2.2.2	Key results	14
2.3	Chemical enrichment	18
2.3.1	Abundance of metals in the gas phase	18
2.3.2	Global metallicity evolution	20
2.3.3	The case of deuterium	21
2.4	Physical properties	22
2.4.1	Theoretical description	22
2.4.2	Observations: methods and results	23
3	Molecular gas	29
3.1	Searches for H ₂ in high-redshift DLAs	30
3.1.1	Using high spectral resolution	30
3.1.2	Medium and low spectral resolution	31
3.2	The high column density gas	32
3.3	Flagging high-metallicity molecular gas using neutral carbon	34
3.3.1	The survey for C I in the SDSS	34
3.3.2	Follow-up with the Very Large Telescope	35
3.4	The H ₂ distribution function from average H ₂ signal in SDSS	38
3.5	Chlorine chemistry	40
3.6	Deuterated molecular hydrogen (HD)	41
3.7	Physical conditions	42

3.7.1	Rotational excitation of H ₂	42
3.7.2	Excitation of C I fine-structure levels	44
3.7.3	Examples: H ₂ at high and low metallicity	44
3.7.4	Analytical expression for the H I surface density of a H ₂ cloud	46
3.7.5	Cloud length-scale	46
3.7.6	Cloud structure	47
3.8	Use of molecular lines for fundamental physics and cosmology	48
3.8.1	Variation of fundamental constants	49
3.8.2	Evolution of the cosmic microwave background temperature	50
4	Dust	54
4.1	Depletion	54
4.2	Reddening of the background quasar light	55
4.3	Dust bias	57
5	Star-formation	59
5.1	Low and intermediate redshifts	59
5.1.1	Connection between metal absorbers and galaxies	59
5.1.2	Searching for 21-cm absorption in or around known galaxies	60
5.2	High redshifts	61
5.2.1	Statistical detection of Ly- α emission	61
5.2.2	Individual direct detection	63
5.2.3	Summarising the DLA-galaxy detections: linking impact parameter to gas column density and metallicity	65
6	Concluding remarks and perspectives	68

1 Introduction

The formation and evolution of galaxies is one of the most fascinating and complex issue of the evolution of the Universe. While the gravitational aspect of structure formation is dictated by dark matter and well captured in high-resolution simulations (e.g. [Klypin et al. 2011](#); [Angulo et al. 2012](#); [Ishiyama 2014](#)), the formation and evolution of galaxies also depends crucially on the properties of the gas in and around them. From its accretion in the form of diffuse gas around the galaxies to its collection in dense self-gravitating clouds that eventually collapse to form stars, the gas undergoes several important transitional processes while its density increases by several orders of magnitude. In addition, the gas integrates the various chemical and physical feedback mechanisms from star-formation activity. Such effects include chemical enrichment of the gas with metals released by stars, production and growth of dust grains, reprocessing and shielding from the ultraviolet radiation field, injection of mechanical energy through stellar winds and supernovae shocks, etc. Part of the gas may also be ejected back into the circum-galactic medium through galactic winds. The new properties of the gas will then determine its evolution and consequently its ability to feed star-formation and so on.

Clearly, this is a complex story in which the gas acts both as a driver and a probe of galaxy evolution. This document summarises my contribution to the deciphering of this story, and aims at presenting simultaneously some conceptual bases as well as analysis techniques developed through the years.

1.1 Gas in nearby galaxies

The gas in and around galaxies can spread a very wide range of physical and chemical properties. Still, we can roughly define phases based on the gas density, temperature and ionisation stage. Warm ionised gas consist of diffuse gas that is photo-ionised by the strong ultraviolet radiation in the vicinity of bright massive stars while a hotter ionised phase is heated by supernova shocks. These mechanisms bring the gas temperature to tens and hundred thousand kelvins, respectively. These ionised phases are not considered further in the present document, where I mostly focus on neutral phases. Neutral phases are of particular interest since they are cooler, denser and contain a large amount of baryons from which stars can be formed. Under typical conditions in the interstellar medium, [Field et al. \(1969\)](#) have shown that the neutral gas tend to segregate into a warm neutral medium (WNM, with temperatures of several thousand kelvins) and a cold neutral medium (CNM, with temperatures of less than a few hundred kelvins). Following works have shown that both these phases can coexist and be in pressure equilibrium with the ionised phase (e.g. [McKee & Ostriker 1977](#)). Finally, a transition from atomic to molecular hydrogen can occur in the CNM, depending on the balance between its formation onto the surface of dust grains, and its dissociation by UV radiation (e.g. [Dalgarno & Stephens 1970](#)).

Before considering gas in the distant Universe, it is worth recalling some major findings from observations in the local Universe. In order to get out of the complexity inherent to the interplay between gas and stars, astronomers first considered the global gas content and star-formation rate at the scale of a galaxy. Remarkably, a neat correlation was found between the disc-averaged surface density of neutral gas (measured from mapping the 21 cm hyperfine-structure emission by neutral hydrogen) and that of star-formation as $\Sigma_{\text{SFR}} \propto \Sigma_{\text{gas}}^{\alpha}$ with $\alpha \approx -1.4$. This relation, known as the Kennicutt-Schmidt (K-S) law (see [Kennicutt & Evans 2012](#)), and valid over several orders of magnitude, has been widely used as a very convenient recipe for star formation in simulations of galaxy formation. However, recent works have emphasized the need to split the interstellar medium (ISM) into its atomic and molecular hydrogen components for a proper modelling of the formation and evolution of galaxies (e.g. [Lagos et al. 2011](#)). Indeed, recent observations of nearby normal galaxies (non-

starburst) have shown that Σ_{SFR} correlates well with the surface density of molecular hydrogen (Σ_{H_2}), but not with that of atomic gas (e.g. [Wong & Blitz 2002](#); [Kennicutt et al. 2007](#); [Bigiel et al. 2008](#); [Leroy et al. 2008](#); [Schruba et al. 2011](#)). In addition, we know that, at least in our Galaxy, stars form in giant molecular clouds which are characterised by parsec scales, i.e. much smaller than the galactic-scales over which the global K-S law applies. This shows that the star formation in nearby galaxies is more coupled with the molecular phase of the interstellar medium than with the total gas surface density, even where the atomic phase makes most of the ISM. This does not explain yet why star formation cares about the chemical state of the gas (i.e. molecular hydrogen vs atomic hydrogen) but at least it tells us that the onset of star-formation must be somehow related to a conversion of atomic to molecular gas.

1.2 Gas in distant galaxies

Modern telescopes have considerably gained in sensitivity and it is now possible to detect the stellar light emitted by galaxies when the Universe was a few percent of its current age and to map the evolution of the cosmic star-formation rate density (Ω_{SFR}). Ω_{SFR} is found to evolve considerably during the Universe's history, increasing during the first 3.5 Gyrs after the Big-Bang, peaking at $z \sim 2$ with a star-formation rate ten times higher than today, and then declining exponentially until $z = 0$ (see e.g. [Madau & Dickinson 2014](#)). Together with the evolution of the global star-formation rate, galaxies also evolve in mass, shape and gas content. Naturally, the distribution, chemical and physical properties of this gas are also expected to evolve. There is therefore no doubt that a complete understanding of the formation and evolution of galaxies cannot rely only on observations in the local Universe and that measurements of the gas at different redshifts, hence at different stages of the evolution of galaxies, is fundamental. This is a main motivation to my work.

Detecting the gas in the distant Universe is actually another challenge. The main technique used to detect and map the atomic gas in nearby galaxies, namely the detection of 21-cm emission from atomic hydrogen, is not sensitive enough to detect gas in the distant Universe with current facilities. Indeed, the farthest 21-cm emissions detected to date have $z < 0.4$ (e.g. [Catinella et al. 2008](#); [Freudling et al. 2011](#); [Fernández et al. 2016](#)) and observing 21-cm emission at higher redshifts will have to await for new very sensitive facilities such as the Square Kilometre Array (SKA). On the other hand, tremendous progress have recently been made thanks to the detection of molecular phases in the sub-mm domain (see review by [Combes 2018](#)), using indirect tracers such as emission from carbon monoxide – since H_2 in typical conditions does not produce any emission line. However the corresponding galaxies first have to be selected through optical surveys and, even with the Atacama Large Millimeter Array (ALMA), only large molecular reservoirs in rather massive galaxies become detectable. Blind searches of molecular lines are in principle possible but remain limited to the population of sub-mm bright galaxies, which has extreme star-forming properties and is not representative of the majority of galaxies.

1.3 Quasars as cosmological probes of the gas

Another approach for studying the gas in the distant Universe is to consider the absorption produced by the gas in the spectra of bright background sources, as illustrated in Fig. 1. Such a technique provides an extremely sensitive way to directly detect the gas, over more than ten orders of magnitude in column densities and without any prior on the galaxy luminosity. In addition, the gas is detected with the same sensitivity over a wide range of redshifts, just limited by the wavelength ranges covered and the availability of background targets. Quasars, which are bright, numerous and emit light over the full electromagnetic spectrum are background targets of choice. Quasars remain the most widely used class of objects for absorption stud-

ies. However, other objects such as Lyman-break galaxies (LBGs) or gamma-ray burst (GRB) afterglows can also be used in principle as background light sources. Galaxies have been little exploited to date because they are generally much fainter than quasars, and because they are not compact. This makes them of less interest to study the local gas conditions (at small scales), but can provide interesting transverse information on the gas distribution over galactic scales (see [Cooke & O’Meara 2015](#)). Since the number density of LBGs on the sky becomes quickly much higher than that of quasars at magnitudes fainter than $g \sim 23$, they will probably be very useful to probe the tomography of the intergalactic medium using future extremely large telescopes. GRB afterglows are compact, bright and can be detected at very high redshift, so that their use as cosmological probe is gaining interest. However, GRB afterglows remain relatively rare events so that it is not possible to compete with the statistics offered by quasars. In addition, the observation of GRB afterglows suffers from the difficulties inherent to transient processes. In the following of this document, I present results derived along quasar lines of sight, but most techniques presented are also applicable to other types of background sources.

The most prominent line observed in absorption is the Lyman- α line of neutral hydrogen (H I) at 1215.67 Å. Since the strength of this line is directly related to the column density of neutral hydrogen along the line of sight, it serves as a very convenient classification for quasar absorption systems. I would like to clarify here some terms used in this document. While the Lyman- α line is produced by *neutral hydrogen*, its detection does not imply at all that the corresponding gas must be neutral. Indeed, the intergalactic gas is mostly ionised but it is the small fraction of *neutral hydrogen atoms* it contains that produces the so-called Lyman- α forest. In other words, H I atoms are here only a tracer of the over-densities of the (ionised) IGM. In turn, by *neutral gas*, one refers to gas that is mostly composed of neutral hydrogen, either in atomic or molecular forms. I will refer here to the first form of neutral gas simply as *atomic gas* and to the later as *molecular gas*. Note that atomic helium is generally not mentioned but implicitly included. For the sake of structuring the present document, atomic gas, molecular gas, dust and star-formation are dealt with in separate sections, although in practice, these galaxy constituents are strongly related to each other and their respective study is not as independent as it may appear from the sectioning.

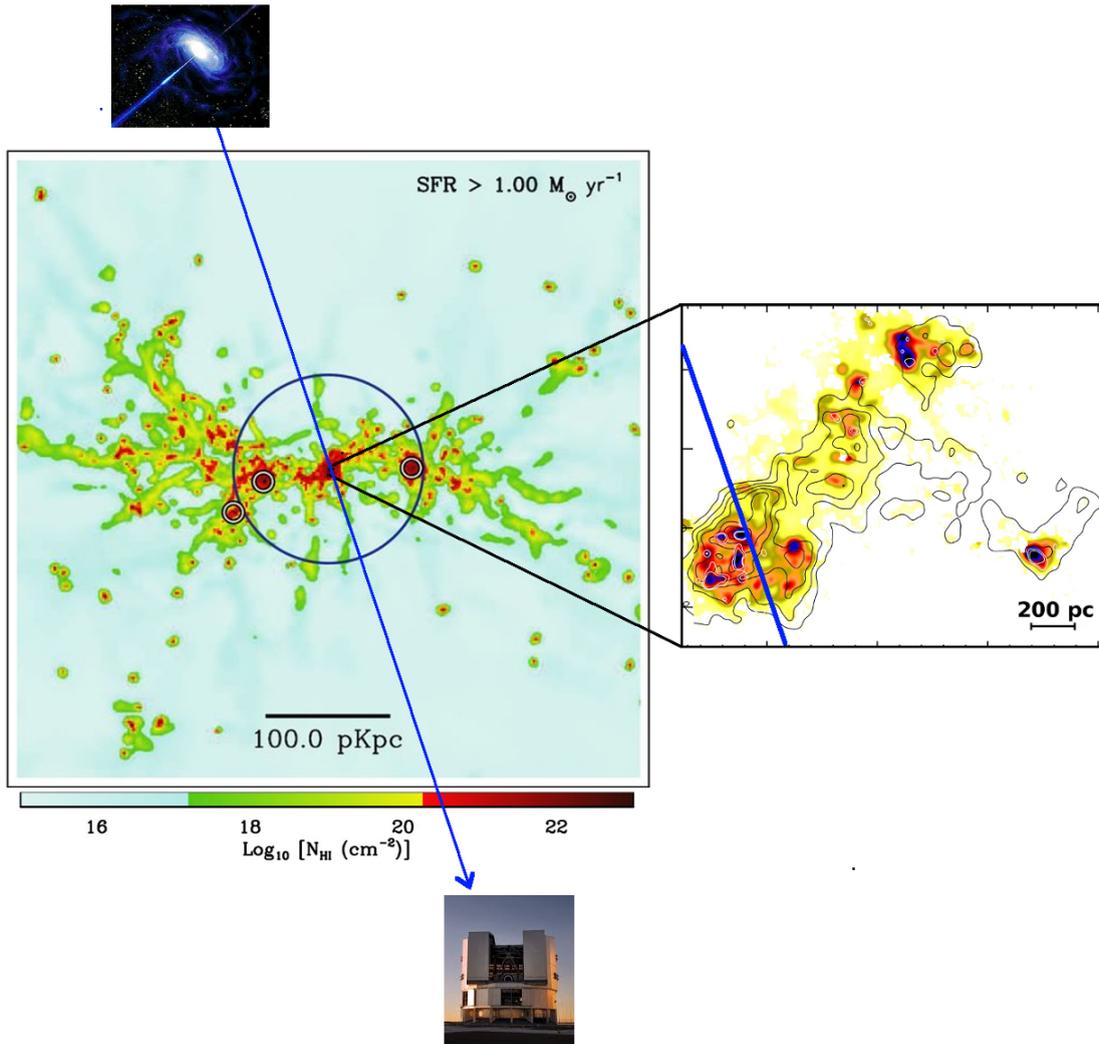


Figure 1: Illustration of quasar absorption spectroscopy to probe gas in and around distant galaxies. Most quasar absorption systems correspond to warm and diffuse gas, that makes out most of the total absorption cross-section (blue and green colours). This gas is also located preferentially outside the main star-forming part of the galaxies. My research has been devoted mostly in characterising the neutral phases (in red), with a recent emphasis on the high column density gas (dark red in the large picture) and the cold molecular gas clumps, found on small scales (illustrated in the small zoomed figure).

2 Atomic gas

When the column density of neutral hydrogen in a given cloud is high enough, ionising photons (i.e. with energies $h\nu > 13.6$ eV) are absorbed by the external layers and cannot penetrate into the cloud. In other words, the gas becomes self-shielded and neutral. The column density above which neutrality occurs depends on the hardness of the ionising source but is expected to take place in the range $N(\text{H I}) \sim 10^{19} - 10^{20} \text{ cm}^{-2}$ (Viegas 1995). Damped Lyman- α systems (DLAs) are the class of quasar absorption systems defined to have $\log N(\text{H I}) \geq 20.3$ (Wolfe et al. 1986) and are therefore predominantly neutral (Wolfe et al. 2005). This definition for DLA is quite fortuitous since the $N(\text{H I})$ -threshold was defined upon the observability of the corresponding Ly- α line in low resolution spectra used at that time ($\log N(\text{H I}) = 20.3$ corresponds to a rest equivalent width of about 10 \AA) and not based on the actual physical properties (neutrality) of the gas. DLA profiles are composed of a saturated region (the "core" of the DLA) and damping wings, which are caused by the natural Lorentzian broadening component of the line. The Doppler (Gaussian) broadening from the thermal and turbulent motion of the atoms is hidden in the saturated profile. An example of a damped Ly- α absorption line is shown on Fig. 2. Because the column density threshold is also similar to the typical column density derived from 21-cm observations of nearby galactic discs, DLAs have soon be considered, somehow abusively, to represent high redshift galactic discs. The picture has now evolved since we know that the atomic gas can be widespread around high redshift galaxies and that galaxies are not necessarily well-formed discs.

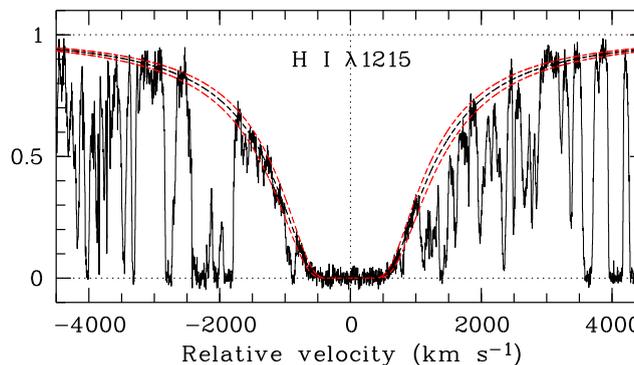


Figure 2: Damped Lyman- α absorption line observed at $z_{\text{abs}} = 1.989$ towards the quasar HE 2318–1107. The derived neutral hydrogen column density is here $\log N(\text{H I}) = 20.68 \pm 0.05$. Note the presence of additional (narrower) absorption lines from the Ly- α forest. [From Noterdaeme et al. 2007a].

2.1 Surveys

The first systematic search for damped Lyman- α , the Lick survey, led to the detection of 15 DLAs along the line of sight to 68 quasars (Wolfe et al. 1986; Turnshek et al. 1989; Wolfe et al. 1993). A number of searches have then followed, each contributing to increase the number of known DLAs and the redshift range probed (e.g. Lanzetta et al. 1991; Wolfe et al. 1995; Lanzetta et al. 1995; Storrie-Lombardi et al. 1996; Storrie-Lombardi & Wolfe 2000; Ellison et al. 2001). The compilations of Curran et al. (2002) and Péroux et al. (2003) listed slightly more than a hundred DLAs at $z > 2$. The advent of the Sloan Digital Sky Survey (see York et al. 2000), which started producing quasar spectra at a high pace, opened the possibility to systematically search for quasar absorption line systems and build large homogeneous samples. With a spectral resolution power of about 2000 and a wavelength coverage starting from about 3800 \AA , SDSS quasar spectra were ideally suited to search for the damped systems at $z > 2.3$. J. Prochaska and his collaborators were the firsts to jump at this chance and published the detection of several hundred DLAs using a semi-automated procedure (Prochaska & Herbert-Fort 2004; Prochaska et al. 2005). In short, their procedure was

based on the identification of spectral regions with observed flux consistent with zero over the typical width of the DLA core. The authors then visually checked the candidates to confirm or not the damped Lyman- α nature of the zero flux regions and fitted manually the lines with Voigt-profiles to derive the systems redshifts and H I column densities.

With the steadily increasing number of available quasar spectra in the SDSS, it became clear that such an approach was reaching its limits. At that time, I was also interested in getting a rough estimate of the metal content of DLAs in order to select targets for high-resolution follow-up with the Very Large Telescope. I therefore developed codes, not only to detect DLA systems, but also to obtain their column density, redshift and metal content in a fully automatic manner. This also included the automatic characterisation of the quasar spectra and the available redshift path to search for DLAs, removing bad spectra and pipeline errors, identifying Lyman-breaks and quasars with broad absorption lines, etc. The originality of my approach was to mimic the criteria used by trained astronomers to recognise damped Lyman- α systems. I proposed to consider not only the saturated core, but also the presence of the damping wings, recognised through correlation analysis, and the association with metal lines to obtain accurate redshift measurements. This resulted in the detection of almost a thousand DLAs in SDSS-DR7 (Noterdaeme et al. 2009b). I continued developing the method over the years, in particular when becoming member of the successors of SDSS-II, the Baryon Oscillation Spectroscopic Survey (SDSS-III/BOSS, Dawson et al. 2013) and the extended-BOSS (SDSS-IV/eBOSS, Dawson et al. 2016). I regularly provide DLA catalogues to the community (see also Noterdaeme et al. 2012b), with a number of systems now reaching several tens thousands at $N(\text{H I}) \geq 10^{20} \text{ cm}^{-2}$.

2.2 Distribution of the atomic gas in the Universe

Since the detection of DLAs depends only on the H I absorption cross-section, the statistics of DLAs along random lines of sight provides fundamental quantities to characterise the overall atomic gas content of the Universe and its evolution with cosmic time. Before discussing the obtained results, I will first present some formalism and definitions of the main quantities that are used to describe the absorbers statistics.

2.2.1 Formalism and definitions

- *Sensitivity function: $g(z)$*

Since to be detected, any absorber must obviously be located between the background source and the observer, the absorber's cosmological redshift is positive and smaller than that of the quasar. However, it is still possible that the observed absorption redshift slightly exceeds these limits due to peculiar motion, in particular of the absorber towards the quasar. This can happen when the gas falls onto the quasar central engine, for example. To avoid including systems that may be physically related to the quasar environment, it is therefore common to restrict the line of sight study to a few thousand km s^{-1} bluewards of the quasar redshift. The minimum accessible redshift is generally set by the data. For example, Ly- α is observable only at wavelengths longer than the atmospheric cutoff from ground based telescopes, which corresponds to $z > 1.6$. A low signal-to-noise ratio or a zero quasar flux due to the presence of a Lyman-break from a system at higher redshift can also limit the ability to detect DLAs in the blue regions of the spectra. All this implies that only a portion $[z_{\min}, z_{\max}]$ along each line of sight can be used to search for a given class of absorption system. The survey size of any survey of quasar absorption systems is then quantified through the sensitivity function $g(z)$, which is simply the number of lines of sight that effectively probe a given redshift.

- *Incidence: $dN/dz, dN/dX$*

The first quantity directly available from any DLA search (or any other class of absorption systems) is the counting statistics, i.e. how many systems are found per unit of pathlength. It is very common to express this as the number of systems per unit redshift, simply called dN/dz . However, because of the expansion of the Universe, even a non-evolving population will have an increasing dN/dz with increasing redshift, so that a more useful way to quantify the pathlength is the absorption distance X , defined as

$$X(z) = \int_0^z (1+z')^2 \frac{H_0}{H(z')} dz', \quad (1)$$

where H_0 is the Hubble constant and $H(z) = H_0 \left[(1+z^3)\Omega_m - (1+z)^2(\Omega_m + \Omega_\Lambda - 1) + \Omega_\Lambda \right]^{1/2}$. A non-evolving population has constant dN/dX with redshift. This quantity is called the incidence and represents the covering fraction of DLAs on the sky for any given redshift. While the DLA incidence is easy to derive, we note that the number of systems increases rapidly with decreasing column density. Consequently, the counted number of DLAs above the threshold is sensitive to any (even-small) systematic error in the $N(\text{HI})$ -measurement. For example, even a slight systematic overestimate of $N(\text{HI})$ can result in the inclusion of a large fraction of sub-DLAs. Therefore, this single number to characterise the DLA population is to be considered with care.

- *Column density distribution function: $f(N, X)(z)$*

Considering now the column densities of each DLA, we define the column density distribution function $f_{\text{HI}}(N, X)$ so that the number of DLAs in the absorption distance interval $(X, X + dX)$ and column density interval $(N, N + dN)$ is given by

$$f_{\text{HI}}(N, X) dN dX \quad (2)$$

- *Cosmological density of atomic gas*

Integrating the column density distribution function over HI column densities above $N_{\text{min}} = 20.3$, we obtain the cosmological density of atomic gas in DLAs

$$\Omega_{\text{g}}^{\text{HI}}(X) dX \equiv \frac{\mu m_{\text{H}} H_0}{c \rho_{\text{c}}} \int_{N_{\text{min}}}^{\infty} N(\text{HI}) f_{\text{HI}}(N, X) dX, \quad (3)$$

where ρ_{c} is the critical mass density of the Universe. We note that the mean molecular mass of the gas, $\mu = 1.3$, has been introduced here to take into account for helium¹. [Wolfe et al. \(2005\)](#) argued that systems with column densities lower than the DLA threshold may not be completely neutral and should not be included when discussing the census of atomic gas. Their contribution to the total atomic gas content can however be estimated from extrapolating the shape of the above integral over bins of column density. We will see that, contrary to the incidence of DLAs, the total amount of neutral gas that DLA contain is less sensitive to N_{min} .

¹It is worth mentioning that not all authors include this constant in their definition, which then corresponds to the cosmological density of atomic hydrogen only.

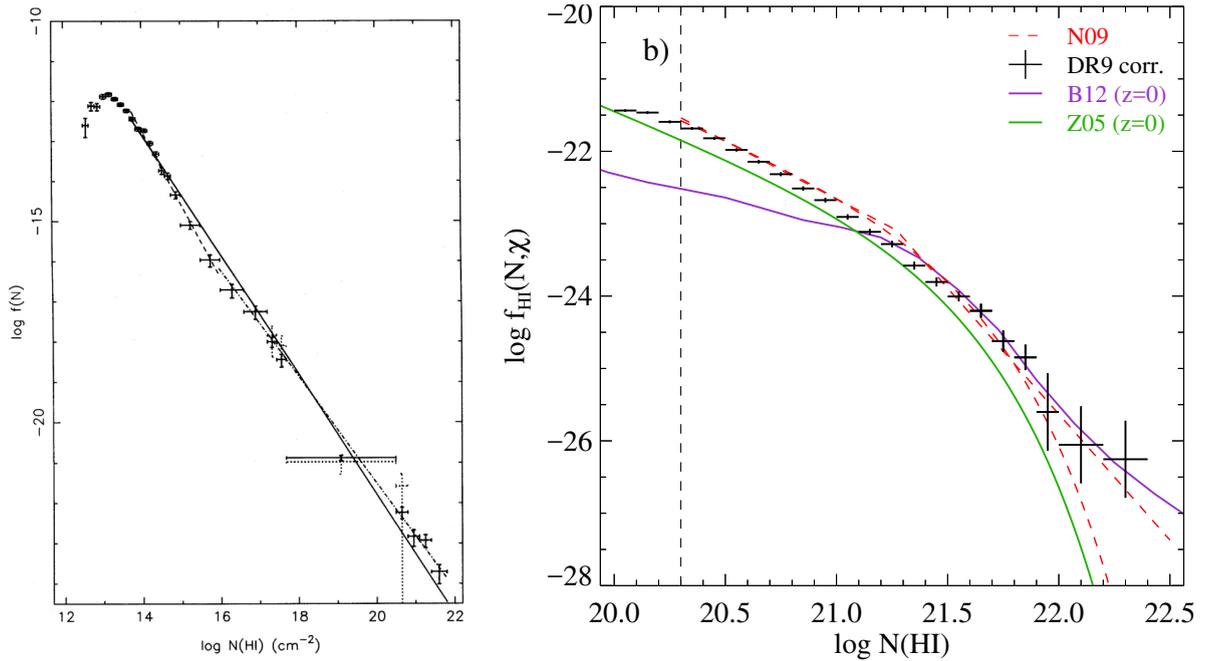


Figure 3: Column density distribution functions at $\langle z \rangle = 2.5$ early derived by [Petitjean et al. \(1993\)](#) over the full range from Ly- α forest to DLAs (left) and that obtained for $\log N(\text{HI}) > 20$ using the SDSS-III ([Noterdaeme et al. 2012b](#), black points in the right panel). Horizontal bars represent the bin over which $f_{\text{HI}}(z)$ is calculated and vertical error bars represent Poisson uncertainty. Note the improvement in the DLA regime obtained thanks to tremendous increase in statistics: the right panel corresponds to the only four points at the bottom right in the left panel. The double power-law and Γ -function fits to the DR7 distribution ([Noterdaeme et al. 2009a](#), $\langle z \rangle = 2.9$) are shown as red dashed lines on the right figure. The f_{HI} at $z = 0$ from [Braun \(2012\)](#), purple) and [Zwaan et al. \(2005\)](#), green) are also shown for comparison.

2.2.2 Key results

The column density distribution function is the quantity that contains most information and from which other quantities such as the cosmological density of neutral gas can be derived. Fig. 3 presents side by side the column density distribution function measured over the full column density range starting from the Ly-alpha forest by [Petitjean et al. \(1993\)](#) and the distribution in the damped Lyman- α regime that I derived from the SDSS (see [Noterdaeme et al. 2009a, 2012b](#) and Table 1). The column density distribution roughly follows a single power law when considering the full column density range from $\log N(\text{HI}) = 12$ to 22, but with very interesting deviations that reflect the different physical regimes. From Fig. 3, the tremendous increase in statistics concerning the neutral gas absorbers is evident. In particular, for this regime, we can immediately see that the distribution function is relatively shallow close to the conventional DLA threshold but steepens at the very high column-density end.

The flattening of the H I distribution function around $\log N(\text{HI}) \sim 20$ results from the change of regime from neutral to ionised gas when going from high to low column densities. This flattening was already noticed by [Péroux et al. \(2003\)](#) but was not clearly apparent in the survey of [Prochaska et al. \(2005\)](#). It is now seen beyond any doubt from my DLA survey in SDSS-III. Not only the statistical errors are very small in this regime owing to the very large number of systems detected but my work also included a global correction for systematics using realistic mock spectra provided by the SDSS collaboration. It is important to quantify the flattening of the distribution function to estimate the contribution of systems below the conventional DLA-cutoff to the total neutral gas content of the Universe, and to demonstrate (rather than postulate) that DLAs do contain the bulk of the H I gas. Indeed, DLAs were long presented as containing the vast majority of the

Table 1: $N(\text{HI})$ distribution function at $\langle z \rangle = 2.5$

$\log N(\text{HI})$	$\log f_{\text{HI}}(N, X)$	$\log f_{\text{HI}}^{\text{corr.}}(N, X)$	$\sigma(\log f_{\text{HI}}(N, X))$
[20.00,20.10[-21.20	-21.44	0.02
[20.10,20.20[-21.37	-21.47	0.02
[20.20,20.30[-21.55	-21.59	0.02
[20.30,20.40[-21.66	-21.68	0.02
[20.40,20.50[-21.81	-21.82	0.02
[20.50,20.60[-21.97	-21.98	0.02
[20.60,20.70[-22.13	-22.14	0.03
[20.70,20.80[-22.30	-22.32	0.03
[20.80,20.90[-22.49	-22.51	0.03
[20.90,21.00[-22.63	-22.67	0.03
[21.00,21.10[-22.85	-22.91	0.04
[21.10,21.20[-23.04	-23.11	0.04
[21.20,21.30[-23.19	-23.28	0.05
[21.30,21.40[-23.46	-23.58	0.06
[21.40,21.50[-23.66	-23.81	0.07
[21.50,21.60[-23.83	-24.01	0.08
[21.60,21.70[-24.20	-24.20	0.08
[21.70,21.80[-24.62	-24.62	0.12
[21.80,21.90[-24.85	-24.85	0.18
[21.90,22.00[-25.60	-25.60	0.53
[22.00,22.20[-26.05	-26.05	0.53
[22.20,22.40[-26.25	-26.25	0.53

Notes: The third column takes into account a correction for systematic effects, estimated from applying the algorithm on mock spectra. The last column corresponds to Poissonian uncertainties.

neutral gas at high redshift (see review by [Wolfe et al. 2005](#)) but this statement relied only on the neutrality threshold argument and was not yet demonstrated from neutral gas census until we constrained the slope of the H I distribution function at both high and low column densities.

To understand this better, Fig. 4 shows another representation of the column density distribution of neutral gas, where the frequency distribution function is integrated over each bin of column density. In other words, the distribution shown on Fig. 4 represents the contribution of DLAs with a given column density to the total atomic gas content in the Universe. Since the amount of neutral gas is the product of the column densities and the incidence of absorbers, the contribution from systems at both ends of the distribution is actually small². In fact, for the total H I content to be finite, the integration of the distribution function overall all column densities must converge. This implies that its slope must be steeper than -2 at the high column density end (as already discussed by [Prochaska et al. 2005](#)) but also shallower than -2 at the low column density end, which is not seen nor discussed by these authors. Fig. 4 does show that DLAs contain the bulk of the atomic gas at high redshift. Extrapolating the slope to lower column densities, we can estimate this contribution to be about 80%, close to what has been estimated in the local Universe from 21-cm studies ([Zwaan et al. 2005](#)), since both frequency distribution have similar shapes. Nevertheless, we found a larger contribution of high column density systems at high redshift. Interestingly, [Braun \(2012\)](#) argued that [Zwaan et al.](#) underestimated the contribution from very high $N(\text{H I})$ systems since these authors did not apply opacity corrections. [Braun \(2012\)](#) derived a distribution (purple line on Fig. 4) which is actually more similar to that seen at high redshift at the high column density end. On the other hand, the work by [Braun](#) is based only on three galaxy maps which may crop part of the extended low column density gas. This would then explain the differences at low $N(\text{H I})$. The neat result is that, trusting the opacity-corrected values from [Braun \(2012\)](#) at high $N(\text{H I})$, and the large statistics of [Zwaan et al. \(2005\)](#) at low $N(\text{H I})$, the shape of the local $f_{\text{HI}}(N, X)$ becomes actually very similar to that seen at high- z , where the above-mentioned issues are avoided thanks to the use of blind quasar absorption spectroscopy. In summary, it is well possible that the shape of the frequency distribution evolves only slightly.

By integrating the column density distribution over the overall range of column densities, we obtain the cosmological density of atomic gas. Figure 5 presents measurements at different redshifts: At $z \sim 2 - 3.5$, the values were obtained directly from SDSS studies. At $z > 3.5$, the Lyman- α forest become very dense and line blending makes the detection and column density measurements difficult at low resolution. Observing $z > 4$ quasars at twice the SDSS resolution ([Guimarães et al. 2009](#)), we showed that Ω_{HI} may actually peak around $z \sim 3.5$ and decrease at higher- z , but the statistics of this sample remain small. At $z < 2$, the Lyman- α lines is located out of the SDSS wavelength range. At $z < 1.6$, the Ly- α line even falls bluewards of the atmospheric cutoff and DLAs can only be detected directly from space observations. Indirect measurements using strong Mg II systems as proxies for DLAs have been used ([Rao et al. 2006](#)) but these observations are challenged by more recent HST/COS observations that favour much lower values ([Neeleman et al. 2015](#)). Finally, some discrepancies are seen between different studies at $z = 0$, which likely arise from the issues discussed above.

Without looking into details (the scale of the figure is linear after all), we can derive broad conclusions: First, atomic gas accounts for about 0.1% of the total mass of the Universe and this amount evolved little over the past 12 Gyr. This contrasts with the global star-formation rate density that changes by a factor 10 over the same timescales (see review by [Madau & Dickinson 2014](#)). Second, the cosmological density of atomic gas is less than the cosmological density of baryons in stars (Ω_{\star}) at $z = 0$. In other words, DLAs do not contain enough atomic gas to account for the amount of baryons in stars at $z = 0$. These results are not in contradiction with the fact that stars form out of the neutral gas, but implies that the neutral gas is just a

²Despite being the most numerous, systems with low $N(\text{H I})$ contribute little to the global H I census. On the other hands, systems with very large $N(\text{H I})$ are very rare and then also contribute little to Ω_{HI} .

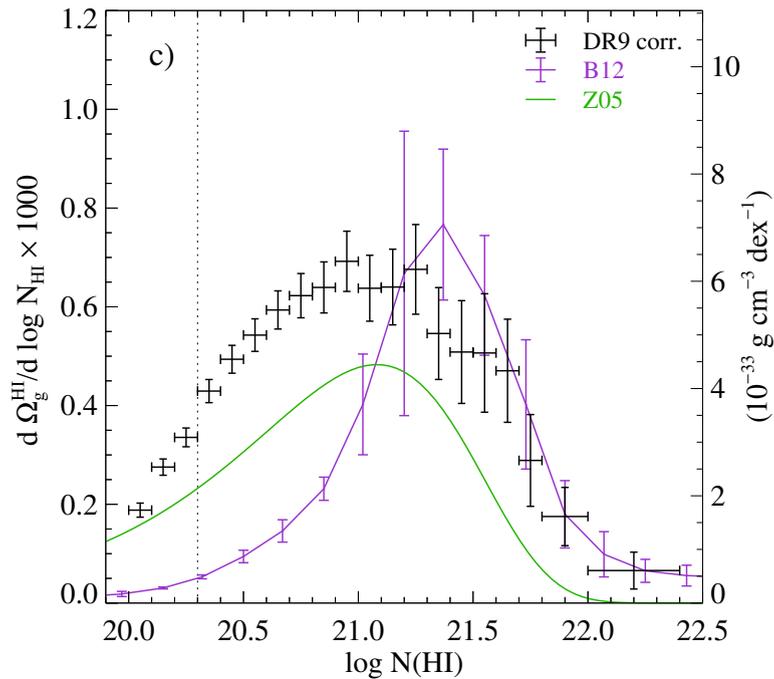


Figure 4: Contribution of absorption systems in a given $N(\text{HI})$ -range to the total mass census of atomic gas. Colours and references are the same as in Fig. 3. [From [Noterdaeme et al. 2012b](#)]

transitional phase that must be constantly replenished. Indeed, the amount of atomic gas in galaxies is set by the competition between its consumption by star-formation and the accretion of baryons from the intergalactic medium. Several works have suggested that the star-formation rate approximately equates the accretion rate, owing to self regulating mechanisms (e.g. [Bouché et al. 2010](#); [Shi et al. 2011](#); [Davé et al. 2012](#); [Lilly et al. 2013](#)): a high star-formation rate will imply galactic winds that will expel part of the gas from the galaxy, hence decreasing star-formation. On the other hand, accretion of fresh gas will trigger star-formation.

The measurements of Ω_{HI} at different redshifts provide now strong constraints that must be reproduced by simulations of galaxy formation and evolution. Further progress on the observational side would require increasing the statistics at $z < 2$ and at $z > 3.5$. This is possible from blind DLA detection using space telescopes for $z < 2$ (although very costly in terms of current telescope time) and medium-resolution spectroscopy for $z > 3.5$ (with e.g. the future Maunakea Spectroscopic Explorer). Meanwhile, thanks to SDSS, the DLA statistics have become enormous at $2 < z < 3.5$ and we have reached a regime where systematics dominate the error budget. This means that any progress in this redshift range is likely to come from comparing different searching algorithms applied to the same data, and better controlling biases due to the quasar selection itself. However, increasing the statistics there as well is actually useful when considering the amount of atomic gas *per* $N(\text{HI})$ bin. This is particularly true for the very high column densities, where the number of known systems remains low. Using the SDSS-III Data Release 11, I have built a sample of about a hundred extremely strong DLAs, with $N(\text{HI}) > 5 \times 10^{21} \text{ cm}^{-2}$ ([Noterdaeme et al. 2014](#)). This constrains the high end of the column density distribution to unprecedented accuracy and up to unexplored column densities. Very high column density absorbers correspond to the regime that is sensitive to both H_2 formation and stellar feedback and hence provide precious constraints for simulations as well ([Altay et al. 2013](#)). We will indeed come back to discussing very high column density systems both in Sect. 3 and Sect. 5.

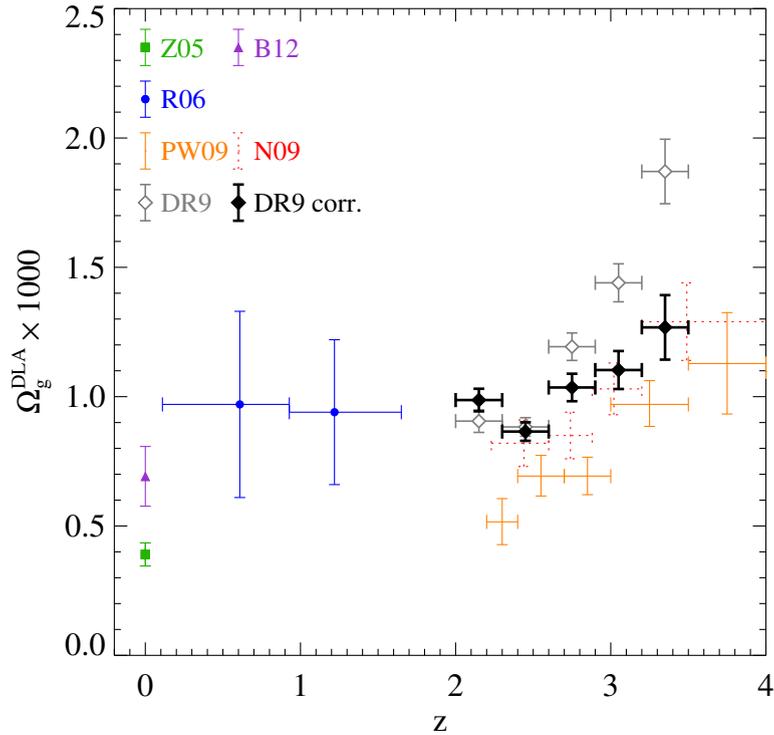


Figure 5: Evolution of the cosmological mass density of atomic gas in DLAs. References are Z05: Zwaan et al. (2005), B12: Braun (2012), R06: Rao et al. (2006), PW09: Prochaska & Wolfe (2009), N09: Noterdaeme et al. (2009a), DR9: Noterdaeme et al. (2012b). Black points include a global correction for systematic effects using mock data.

2.3 Chemical enrichment

The presence of different heavy elements in the atomic gas results from the enrichment by successive generation of stars so that the measurement of their abundances probes the star-formation history. Indeed, measuring the metallicities in galaxies at all redshifts is a major topic in astrophysics and we will see here that DLAs provide a very useful tool to this end. Metal absorption lines are ubiquitous in DLAs with a wealth of metal species detected in different ionisation stages (e.g. neutral oxygen and nitrogen, singly-ionised zinc, iron and silicon as well as higher ionisation species such as C IV or Al III). This is illustrated in Fig. 6 which shows a composite DLA spectrum at $z \sim 2 - 3$ that we built by averaging a large number of SDSS spectra shifted to the DLA rest-frame (Mas-Ribas et al. 2017).

2.3.1 Abundance of metals in the gas phase

Abundance measurements in the neutral gas are robust in absorption for two reasons. First, contrary to emission-line-based measurements, the column density of a given species is measured in absorption independently of the gas physical state such as density or ionisation. Indeed, the direct observable in absorption is the optical depth $\tau(\lambda)$, which depends only on the column density N , the atomic parameters and the line Doppler parameter b . In fact, in both the optically thin and the damped regimes, the line equivalent width does not depend on b at all and the measurement is greatly facilitated. In the intermediate (logarithmic) regime, the dependence of the line equivalent width on b can be problematic in particular since the typical turbulent broadening are of a few km/s, i.e. smaller than the spectral resolution of most instruments. However, several lines spanning a range of oscillator strengths are generally available for a given species, so that it is possible to actually derive both N and b simultaneously. Furthermore, different species are found in the neutral gas

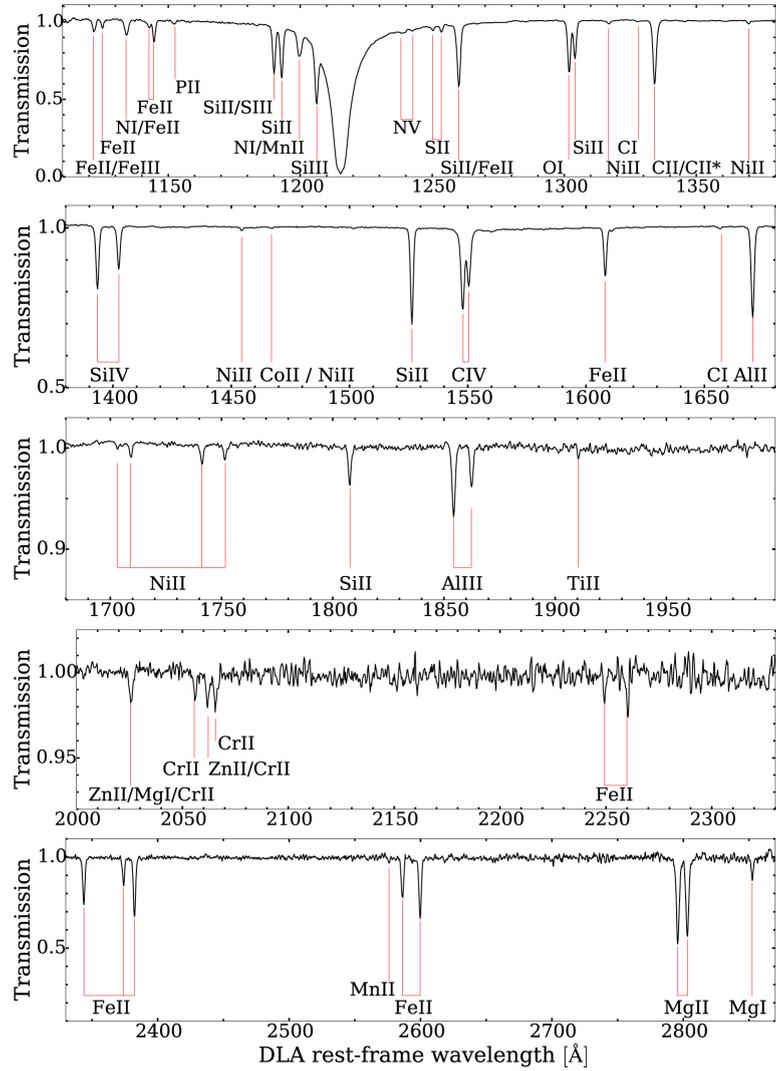


Figure 6: Composite transmission spectrum from the SDSS DLA catalogue. Labels under the red vertical lines denote the species. Each panel has a different scaling for the transmission, selected to enhance the visibility of the lines. [From Mas-Ribas et al. 2017]

and it is reasonable to assume that they share a common kinematic pattern, i.e. same velocity components and Doppler parameters³. This means that the simultaneous analysis of different metal lines allows to constrain the abundance of each species better.

The second reason why abundance measurements in absorption are robust in the neutral gas is that most atoms of a given element are found in a same ionisation state, which depends on whether their first ionisation potential is above or below the hydrogen ionisation energy (13.6 eV). For example, iron, sulphur, silicon, zinc are found almost exclusively in respectively Fe II, S II, Si II, Zn II while nitrogen or oxygen have high first ionisation potential and are found in neutral form (N I, O I). The relative abundance of a given element in its main ionisation stage with respect to neutral hydrogen is then a very good approximation of its true abundance relative to hydrogen. In other words, ionisation corrections should be very small in the neutral gas (e.g. Péroux et al. 2007). We note that while absorption lines from highly-ionised metals (e.g. C IV, Si IV) are also seen associated to DLAs, these likely arise from ionised gas which is not co-spatial with the atomic gas, as also suggested by their different observed kinematics (see Fox et al. 2007).

2.3.2 Global metallicity evolution

Since DLAs are selected irrespectively of the properties of the associated object, measuring the metal abundances in a large sample of systems over a wide range of redshifts currently provides the best picture of the chemical enrichment of atomic gas in the Universe. It is therefore not surprising that measuring DLA metallicities constitutes a main topic in the field (e.g. Lu et al. 1996; Kulkarni & Fall 2002; Prochaska et al. 2003; Ellison et al. 2005; Rao et al. 2006; Dessauges-Zavadsky et al. 2006; Péroux et al. 2007; Rafelski et al. 2012, 2014; De Cia et al. 2018, among many other works). Metallicity measurements are generally based on the abundance of zinc or sulphur, which are volatile species, while most other species tend to deplete onto dust grains. As a consequence, their abundance measured *in the gas-phase*, although robustly determined, is a lower limit to their *total* abundance, which is the quantity of interest when discussing global metallicity evolution. Unfortunately, the abundance of zinc and sulphur is not always measurable since S II lines are generally located in the Ly- α forest while Zn II lines, with $\lambda = 2026, 2062$, are rapidly redshifted in regions crowded with sky lines, or not covered by the spectroscope. Rafelski et al. (2014) have therefore mostly used silicon and iron to estimate DLA metallicities at the highest redshifts. However, this species are depleted even at low metallicity, meaning that any conclusion based on such measurements remains questionable. Actually, even volatile elements can present some level of depletion at high metallicity. Fig. 7 shows a recent compilation of measurements performed at high spectral resolution and where corrections for depletion onto dust grains have been applied (De Cia et al. 2018).

Having said this, while the sample size has grown constantly through the years and measurements getting more and more accurate, the main results remain qualitatively unchanged: First, metals are detected in all DLAs, with the most extreme metal poor systems still having metallicity of several thousandths of the Solar value. This means that for the redshift range probed, there has always been some enrichment of the gas by stars. Second, the mean cosmic metallicity increases with cosmic time, as expected if successive generation of stars enrich more and more the gas. Third, there is a large dispersion between the values in different DLAs which shows that DLA probe a wide range of star-formation histories. All this clearly indicates that DLAs are associated with galaxies, although it does not tell us (yet) where this atomic gas is found with respect to the star-forming regions.

³We note here that the observed Doppler parameter is the quadratic sum of a turbulent and a thermal component, the former generally dominating the later in the typical conditions in the neutral gas and for heavy species. We will discuss this further in Sec. 2.4.2.

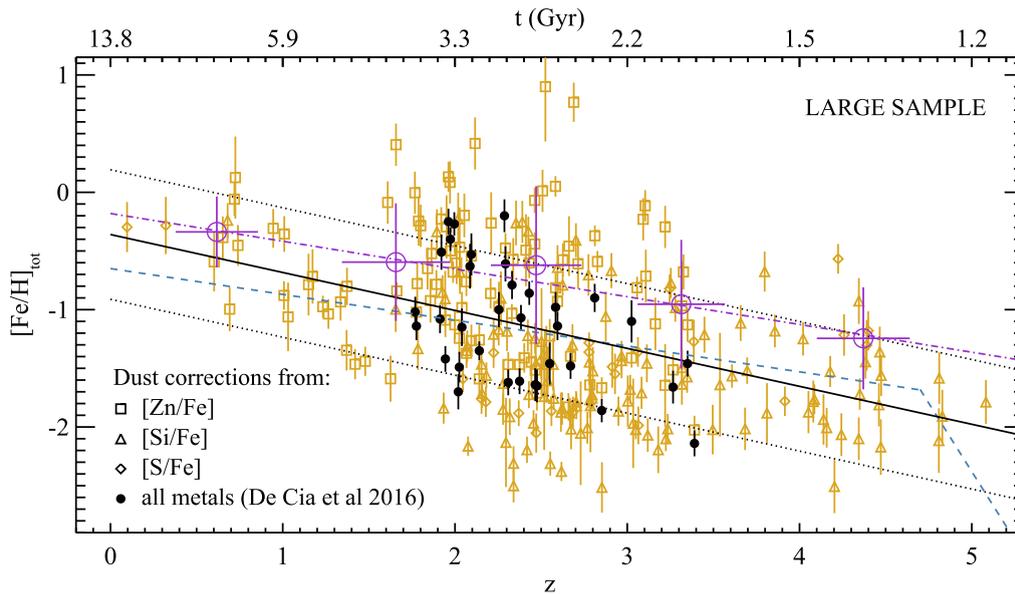


Figure 7: Evolution of the total (i.e. dust-corrected) metallicity in DLAs. The purple points show the $N(\text{H I})$ -weighted average in different redshift bins, which indicates an almost linear evolution with redshift (purple dashed-dotted line). The blue dashed line shows the result without dust-correction by Rafelski et al. (2014) who observed a possible rapid drop at $z > 4.7$. Overall, De Cia et al. (2018) found metal abundances 0.3-0.5 dex higher than previously thought. [From De Cia et al. 2018]

2.3.3 The case of deuterium

Among all species that can be found in the atomic gas, deuterium (^2H or D) is a special case since it is not produced by stellar nucleosynthesis. On the contrary, deuterium is produced during the Big-Bang and subsequently destroyed in stars interiors. Since the primordial abundance of deuterium directly depends on the baryon-to-photon ratio η , the (D/H) ratio in pristine gas provides a baryometer of choice (e.g. Steigman et al. 2007). However, measuring the abundance of deuterium is observationally very challenging. First, its abundance is about 10^{-5} that of hydrogen and second, D I lines are located only about 80 km s^{-1} from the much stronger H I lines. This explains why, despite the huge number of DLAs currently identified, there has been only a few measurements of deuterium. Ryan Cooke and collaborators have focused on getting accurate measurements in the most metal poor systems, in which the gas has been little processed by stars, with the scope of measuring the primordial abundance and constraining Ω_b (Cooke et al. 2014). Another advantage of low-metallicity systems is that the kinematic profile are generally simple with only a few components (see Ledoux et al. 2006a). However, given the paucity of such measurements, we cannot afford to disregard measurements at different metallicities. Furthermore, models predict the dispersion and slow decrease of deuterium abundance as a function of gas enrichment (Dvorkin et al. 2016), so that the full metallicity range actually becomes of interest, not only to globally constrain the primordial abundance, but also to understand deuterium astration by stars.

Archival high-resolution spectroscopic data can be a precious source for measurements at no additional observational cost. Vincent Dumont, who was a master student under supervision by Sebastian Lopéz and myself, helped us to skim the full set of archival VLT/UVES data available at that time, which resulted in a new detection of deuterium at high redshift ($z_{\text{abs}} = 2.621$ towards the quasar CTQ 247). While the $(\text{D I}/\text{H I})$ was only obtained after assuming a constant metallicity throughout the system, the (D/O) ratio is derived very accurately (Noterdaeme et al. 2012a) thanks to charge exchange that maintain this ratio equal to the measured $(\text{D I}/\text{O I})$. This ratio is also useful to constrain evolution models since the abundance of oxygen is expected

to increase with cosmic time, while that of deuterium decreases (see also Hébrard & Moos 2003 for similar measurements in our own Galaxy). In addition, as we will see in Sect. 2.4.2, deuterium is very helpful to directly measure the temperature of the atomic gas.

2.4 Physical properties

In this section, I discuss the physical conditions in the atomic gas. It is known that under the pressure exercised by the total weight of material in the galaxy, the atomic gas tends to segregate into a warm neutral medium (WNM, with high temperatures, $T \sim 10^4$ K, and low densities, $n_{\text{H}} \sim 0.1 \text{ cm}^{-3}$, and a cold neutral medium (CNM, with low temperatures, $T < 500$ K, and higher densities, $n_{\text{H}} \sim 100 \text{ cm}^{-3}$). While stars are unlikely to form out of warm gas, the cold phase corresponds to gas clouds that can potentially further cool down and compress until a final gravitational collapse. Determining the temperature and density of the atomic gas and in particular the relative fraction of CNM and WNM is therefore a major topic in DLA studies.

2.4.1 Theoretical description

A first, popular model for the physical state of the atomic gas is the two-phases model originally developed by Field et al. (1969). The idea is to derive the gas equation of state assuming thermal equilibrium, i.e. a balance between the cooling and heating rates. It is not my purpose here to detail all the different heating and cooling mechanisms, which are described in several works (e.g. Wolfire et al. 1995), but to present the main processes that dictate the behaviour of the gas.

In their early model, Field et al. considered heating of the gas by cosmic rays, but subsequent works based on the observation of molecules sensitive to the ionization rate suggested that the cosmic ray ionisation rate was lower than assumed and may not be the dominant heating source. Wolfire et al. (1995) have shown that the primarily heating process is actually photo-electric effect onto dust grains: UV photons are absorbed by dust grains which leads to the ejection of electrons. After taking into account the energy consumed to liberate the electron, there is about 1 eV of energy per ejected electron that is available for gas heating. The heating rate therefore depends on the product of the intensity of the UV field and the dust-to-gas ratio.

Cooling of the gas is dominated by collisional excitation of fine-structure levels of different species that liberate the energy through emission lines, the most important of which being $[\text{C II}]\lambda 158\mu\text{m}$. Indeed, carbon has a high abundance and a low ionisation potential (11.26 eV, i.e. below the 13.6 eV of hydrogen ionisation). Moreover, the fine-structure transition of ionised carbon is characterised by a low temperature and its critical density for excitation through collisions is also low (see e.g. Lagache et al. 2018). $[\text{O I}]$ emission at $145\mu\text{m}$ is also an important coolant, in particular at low densities. One can see that the cooling of the gas thus depends on the product of the density of metal species and the density of collision partners. In other words, at the first order, the cooling is proportional to $n^2 Z$, where Z is the gas metallicity.

The different cooling and heating mechanisms are shown as a function of the gas density in the left panel of Figure 8 for a standard model by Wolfire et al. (1995). Solving the thermal equilibrium at constant density n gives the phase diagram shown on the right panel of this figure. Under a low pressure (here $(P/k) < (P_{\text{min}}/k) \sim 10^3 \text{ K cm}^{-3}$), the gas can only have a low density and a high temperature. Under a high pressure (here $(P/k) > (P_{\text{max}}/k) \sim 3600 \text{ K cm}^{-3}$), the gas is cold and of high density. A stable two-phase medium, consisting of a mixture of both CNM and WNM exists for $P_{\text{min}} < P < P_{\text{max}}$ ⁴. This happens to be the case for the typical pressures in the discs of galaxies. We note that several works indicate the existence of gas with

⁴A third solution, in-between the two inflexion points of the phase diagram (at around $n \sim 1 \text{ cm}^{-3}$), has $dP/dn < 0$ and is hence unstable.

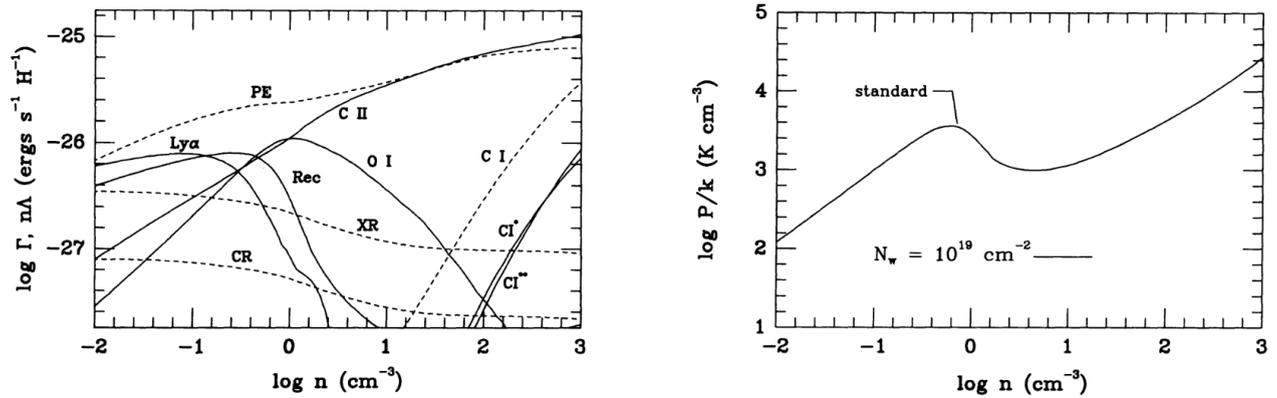


Figure 8: *Left:* Heating rates (dashed lines) and cooling rates (solid lines) for different mechanisms as a function of the gas density. The photo-electric effect on dust grains ("PE") dominates the heating at all densities, followed by soft X-ray ionisation ("XR") and cosmic rays ("CR"). Cooling is mainly due to emission lines, with the most important coolant at $\log n > 0$ being [C II]. *Right:* Phase diagram for a canonical "standard" model. Note there are three solutions for pressures of a few 10^3 K cm⁻³, but only two (WNM and CNM), with $dP/dn > 0$, are stable. [From Wolfire et al. 1995]

intermediate temperature which is maintained in local pressure equilibrium by turbulence. The two-phase model remains however the main paradigm and illustrates how the physical state of the gas depends on the pressure set by the environment.

2.4.2 Observations: methods and results

In this section, we present different methods to determine the physical state of the atomic gas from the observations of quasar absorption lines.

Thermal broadening

The thermal (microscopic) agitation of the atoms directly translates into a broadening of the lines, which is given by the thermal Doppler parameter

$$b_T = \sqrt{2k_B T/m}, \quad (4)$$

where m is the mass of the considered atom, k_B is the Boltzmann constant and T is the gas temperature. However, macroscopic motion of the gas (i.e. turbulence) is also at play so that the observed Doppler parameter is actually given by $b^2 = b_T^2 + b_{\text{turb}}^2$ where b_{turb} is the turbulent component. While there is no way to disentangle thermal from turbulent broadening using a single species, it is possible to derive the relative contribution of turbulence and temperature broadening by comparing the observed Doppler parameters of species with different masses, assuming they are found in the same gas. In practice however, b_{turb} typically reaches several kilometres per second and generally dominates over b_T for heavy species, even in the warm neutral medium. For example, the thermal broadening of iron (resp. silicon) lines is still only 1.7 km s⁻¹ (resp. 2.4 km s⁻¹) at 10^4 K. For a turbulent broadening of 5 km s⁻¹, the actual differential observed broadening between silicon and iron would then be less than 0.3 km s⁻¹. In fact, it is a common simplification to assume a same b for all metal species in a given gas phase.

Nevertheless, from a careful analysis of oxygen, silicon and iron lines in a low-metallicity DLA system with simple kinematic profile, Carswell et al. (2012) were able to measure the slight difference in line widths of the species and found the gas temperature to be $T = 12000 \pm 3900$ K. This corresponds to a warm neutral medium.

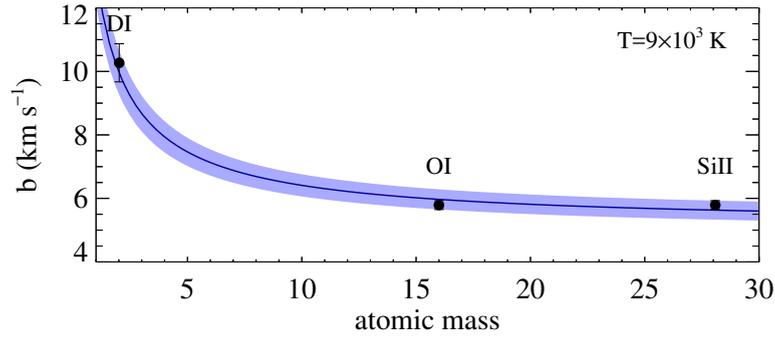


Figure 9: Measurement of the gas temperature by disentangling the thermal and turbulent broadening of different lines. The total (observed) Doppler parameter asymptotically equates the turbulent component at high masses, when the contribution of the thermal component clearly dominates for deuterium. The solid line with associated error as shaded blue region is a fit to the observed Doppler parameters using $b = \sqrt{b_{\text{turb}}^2 + b_T^2}$, with $b_T^2 = 2k_B T/m$, indicating $T \approx 9000$ K. [From [Noterdaeme et al. 2012a](#)]

Because of the inverse dependence on mass, determining the thermal broadening should be strongly eased by observing lighter species. Unfortunately, hydrogen lines are saturated in the atomic gas, meaning that the main observed lines are damped and insensitive to the Doppler parameter. In addition, even for the weakest Lyman lines, it is very difficult to decompose the profile into individual velocity components. Deuterium comes here to the rescue. Not only it is very light so that its thermal broadening is expected to be of the order of 1 and 10 km s⁻¹ (for cold and warm phases, respectively), but its low abundance makes the component identification and the Doppler parameter measurement easy – provided it is not lost within the absorption profile of the associated H I lines. Figure 9 shows a comparison between the observed Doppler parameters for neutral deuterium, neutral oxygen and ionised silicon in the system towards the quasar CTQ 247 ([Noterdaeme et al. 2012a](#)). We note here that oxygen has almost exactly the same ionisation potential as hydrogen and favourable charge exchange reactions make the species tied together, meaning that they trace the same gas. We measured the temperature of the gas to be $T = 8800 \pm 1500$ K, equating the canonical value for the WNM phase ([Wolfire et al. 1995](#)). This is probably the most direct and accurate measurement of the gas temperature in a WNM at high- z so far.

Interestingly enough, we note that taking into account the thermal broadening of the lines can also be important for measuring column density of ionised carbon, in particular since only a single line is observable in DLAs and is generally in the non-linear part of the curve of growth in DLAs (in most of the cases, it is even completely saturated and its column density is not measurable at all). In [Dutta et al. \(2014\)](#), we showed that the carbon abundance measurement by [Cooke et al. \(2011\)](#) at $z_{\text{abs}} = 2.34$ towards J0035–0918 was overestimated by about a factor of ten because of this. The conclusion that this extremely metal-poor DLA is carbon-enhanced compared to iron still stands but it is not as exceptional as claimed by these authors and its interpretation becomes easier. Indeed, increased carbon abundance at low-metallicity is not seen in models of Population II nucleosynthesis and would require either high carbon production by the first generation Population III stars or increased carbon yield from rapidly rotating low-metallicity Population II stars.

21-cm absorption

The hyperfine transition of atomic hydrogen at 21-cm provides a very powerful way to detect the cold gas and derive the CNM fraction since the 21-cm line is sensitive to both the H I column density and the thermal state of the gas ([Kulkarni & Heiles 1988](#)). The 21-cm optical depth of a single homogeneous cloud is related to the spin temperature (T_s) and the H I column density by

$$N(\text{HI}) = 1.823 \times 10^{18} T_s \int \tau(\nu) d\nu \text{ cm}^{-2}. \quad (5)$$

This means that the optical depth is about a factor of hundred larger for cold gas than it is for warm gas at a given column density. In fact, the sole detection of 21-cm absorption is generally enough to be considered as the detection of CNM. For multiple HI clouds along a line of sight, equation 5 remains unchanged except that the spin temperature becomes a column-density-weighted harmonic mean temperature –denoted $\langle T_s \rangle$ – of the different phases that contribute to the observed $N(\text{HI})$. If the HI can be independently determined (e.g. from the Lyman- α absorption line), then HI 21-cm absorption spectroscopy yields $\langle T_s \rangle$. For given spin temperatures of the cold and warm phases, this allows to determine the relative fraction of gas in CNM and WNM phases as

$$f(\text{CNM}) = \frac{\frac{1}{\langle T_s \rangle} - \frac{1}{T_s^W}}{\frac{1}{T_s^C} - \frac{1}{T_s^W}} \approx T_s^C \left(\frac{1}{\langle T_s \rangle} - \frac{1}{T_s^W} \right) \text{ for } T_s^W \gg T_s^C \quad (6)$$

where T_s^C and T_s^W are the spin temperature for the cold and warm phases, respectively. The spin temperature of cold gas can sometimes be estimated from the velocity width of individual components but it is already very instructive to use fiducial values for the CNM and WNM temperatures. For many years, the number of DLAs known towards radio-loud quasars remained very limited, and the poor low-frequency coverage of radio telescopes together with radio frequency interference (RFI) resulted in only a few 21-cm detections and limits in high- z DLAs (e.g. Carilli et al. 1996; Kanekar & Chengalur 2003). It was then hard to conclude whether the paucity of detection was real or only due to observational limitations. Several strategies have been developed in order to address fill this gap. Our team used several facilities such as the Giant Metrewave Radio Telescope (GMRT), the WSRT and the the Green Bank Telescope (GBT) to search for 21-cm absorption at both intermediate redshifts ($z \sim 0.5$ -1) and high redshifts ($z > 2$).

At intermediate redshifts, the Lyman- α line is not observable from ground-based telescopes, so that DLAs have to be pre-selected by other means. Rao et al. (2006) have shown that the Mg II doublet ($\lambda\lambda 2796, 2803$) is very useful to this end, since the probability for a given Mg II system to be a bona-fide DLA becomes high when these lines are strong. I have therefore developed tools to automatically search for Mg II doublets in SDSS spectra and measure accurately the redshift and equivalent width of the different metal lines, which we then used to select the systems to be observed at 21-cm. Through the years, we have increased significantly the number of 21-cm detections and obtained stringent limits for a large number of systems (Gupta et al. 2009). While the CNM fraction cannot be derived without a measurement of the HI column density, the detection statistics is very instructive on its own, and by comparing the incidence of 21-cm to that of DLAs, we can statistically estimate the evolution of cold gas fraction. We have also studied the detectability of 21-cm as a function of the properties of the UV lines (Mg II doublet ratio, Mg I and Fe II equivalent widths) as well as a function of the dust content. This lead us to define optimised selection strategies (Gupta et al. 2012). As expected, the detection probability is increased when the metal lines indicate high HI column densities, and high metal (dust) content. High HI columns will not only increase the optical depth of the 21-cm line for a given temperature, but the cold gas fraction in metal-rich systems could be enhanced. We also found that the number per unit redshift of 21-cm absorbers is almost constant over $0.5 < z < 1.5$, i.e. 30% of the total age of the Universe. This is intriguing since stellar feedback is expected to diminish the CNM filling factor. We concluded that large systematic surveys of 21-cm are necessary to provide a complete view of the evolution of cold gas in galaxies. We are currently working towards this goal with the MeerKAT Absorption Line Survey, during which we will observe about a thousand lines of sight selected without any prior on the presence of absorption systems and where dust biasing is avoided by using IR+radio selection only (see Sect. 4.3).

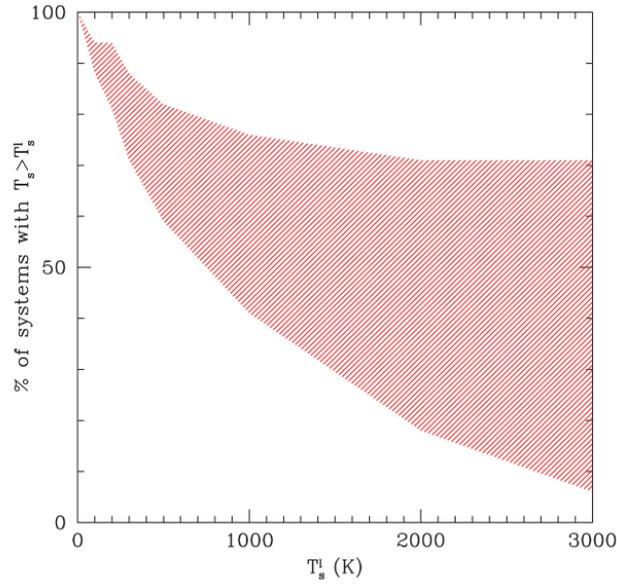


Figure 10: Allowed range on the fraction of high- z DLAs with spin temperature $\langle T_s \rangle$ higher than a limiting value (T_s^l) as a function of this value. [From [Srianand et al. 2012](#)]

In order to interpret correctly 21-cm observations, one should also consider the covering fraction of the background source. If the radio continuum source is more extended than the absorbing gas, which is expected for example in case of radio jets, then the observed optical depth will be decreased by the fraction of radio emission that is not covered by the absorber. Eq. 5 must then include a correction for covering fraction f_c :

$$N(\text{HI}) = 1.823 \times 10^{18} \frac{\langle T_s \rangle}{f_c} \int \tau(\nu) d\nu \text{ cm}^{-2} \quad (7)$$

This means that the interpretation of 21-cm detection statistics can actually be degenerated with the coverage of the background radio source by the (warm+cold) atomic gas. This motivated [Curran \(2010\)](#) to seek for possible geometrical effects in the 21-cm statistics, with the idea that absorbers located closer to the observer should cover statistically more the background source than those located closer to this source. Such a purely geometrical effect seems however not to be dominant compared to the effect of structures both in the radio source and in the gas absorber. Indeed, we have used milliarcsec imaging with the Very Long Baseline Array to study the extent and morphology of the radio sources ([Gupta et al. 2012](#)). We showed that the detection rate can be underestimated by up to a factor of two if 21-cm optical depths are not corrected for the partial coverage estimated using milliarcsecond scale maps.

At high redshift, we selected the targets by cross-matching my SDSS DLA catalogue with catalogues of bright radio quasars. We then performed observations with the GMRT and the GBT, after carefully remeasuring the redshift and column density of the absorbers ([Srianand et al. 2012](#)). In that work, we presented the 21-cm observations of 10 intervening DLAs at $z > 2$ from which we detected 21-cm absorption in a single system but obtained stringent limits for the remaining ones. This sets the fraction of cold gas absorbers to about 10%. We combined our data with measurements from the literature (28 DLAs) and derived the constraints on the harmonic spin temperatures following Eq. 7. The fraction of DLAs having spin temperature more than a given limit is shown on Fig. 10. This figure shows that high spin temperatures $\langle T_s \rangle > 700$ K are found in at 50-80% of the DLAs. Assuming a conservative $T_s^C = 200$ K and $T_s^W \sim 10^4$ K, this implies that the CNM fraction is less than 27% for most of the DLAs. Finally, we found that the fraction of CNM in DLAs tend to increase with increasing $N(\text{HI})$ and metallicity. This is expected since a high metal abundance implies a high

cooling rate, but also, in combination with large gas column densities, a better dust shielding from the UV radiation and hence a decreased heating rate. Large H I column densities are also more likely to be found at small galacto-centric distances (Sect. 5), where the pressure can more easily reach the minimum value for the mixture of CNM and WNM to co-exist.

Excitation of metal fine-structure levels

It has long been recognised that the excitation of fine-structure levels of different species can be used as an indicator of the physical conditions in the gas (Bahcall & Wolf 1968; Smeding & Pottasch 1979). A method to derive the physical conditions, developed by Wolfe et al. (2004), is based on the cooling of the gas through the fine-structure [C II] emission. In its excited fine-structure state, ionised carbon produces an electronic absorption line at 1335.7 Å from which its column density can be measured. The cooling rate l_c through [C II] $\lambda 158\mu\text{m}$ emission can then be estimated following Pottasch et al. (1979) as

$$l_c = \frac{N(\text{C II}^*)h\nu_{ul}A_{ul}}{N(\text{H I})} \text{ erg s}^{-1} \text{ per H atom}, \quad (8)$$

where A_{ul} is the coefficient for spontaneous emission of the $^2P_{3/2} \rightarrow ^2P_{1/2}$ transition and $h\nu_{ul}$ the corresponding energy. Assuming that the [C II] emission dominates the overall cooling rate and that the heating rate is dominated by photo-electric effect on dust grains (and hence proportional to the product of the intensity of the UV field and the abundance of dust), it is possible to estimate the ambient UV field after deriving the abundance of dust from the depletion of refractory elements (Sect. 4). Wolfe et al. (2003) argued that the star-formation rate per unit area would be too large if DLAs were solely composed of WNM (where the cooling is not dominated by [C II] emission anymore) and that a large fraction of the C II* absorption in DLAs arise from the CNM gas (but see however Srianand et al. 2005).

Another method is to consider the population ratio of a metal species in an excited level to its ground state. One advantage of this method is that we do not need to care of depletion, nucleosynthetic effects or ionisation corrections since these are the same for the excited and the ground-state of the given species. Their population ratio depends solely on the physical conditions. A detailed explanation on how to solve the statistical equilibrium is presented in Silva & Viegas (2002). In a steady-state regime, the sum of all processes populating a given level is balanced by the sum over all processes depopulating it. The ratio of excited (2) to ground state level (1) can then be written as

$$\frac{n_2}{n_1} = \frac{B_{12}u_{\nu_{12}} + \Gamma_{12} + \sum_k n_k \gamma_{12}^k}{A_{21} + B_{21}u_{\nu_{12}} + \Gamma_{21} + \sum_k n_k \gamma_{21}^k} \quad (9)$$

where A_{21} , B_{12} and B_{21} are the Einstein coefficients for the given transitions (spontaneous emission, photo-absorption and stimulated emission). $u_{\nu_{12}}$ is the energy density of the radiation field at the frequency of the transition ν_{12} . This is generally dominated by the cosmic microwave background radiation, which can be calculated using standard cosmology⁵. Γ_{12} and Γ_{21} are the fluorescent rates. Finally, the summation describes the (de)excitations with collision partners k ($= e^-, \text{H}^0, \text{H}_2, \text{He}^0, \dots$) with densities n_k and collision rates γ^k . These are related to each other through $\gamma_{12}^k/\gamma_{21}^k \propto e^{-k_B T_{12}/k_B T}$ (Spitzer 1978) where $k_B T_{12}$ is the energy of the excited level to the ground level and T is the gas temperature.

Solving Eq. 9 allows one to determine the physical conditions in the gas. It also happens that several terms can

⁵Conversely, the CMB temperature can be actually inferred from the excitation of the different species, when other processes are less important or well constrained.

be ignored in this equation, depending on which is the dominant process. For example, if collisions dominate, then the fine-structure ratios will allow one to infer the density of the gas. Following a work by [Howk et al. \(2005\)](#), [Neeleman et al. \(2015\)](#) have recently considered the excitation of C II and Si II in a sample of DLAs and derived constraints on the density and temperature assuming the fluorescence rates to be negligible and considering collisions with electron, protons and atomic hydrogen only. These authors conclude that about 5% of DLAs have the bulk of the atomic gas in a CNM phase.

We can again understand these results within the theoretical model presented in Sect. 2.4.1: when thermal pressure excess a minimum value, turbulent pressure fluctuations can produce cold gas that is thermally stable. In galactic discs, the pressure is a function of the galacto-centric radius so that cold gas is found within a certain radius, while at larger distances, only warm gas exists ([Wolfire et al. 2003](#)). At high redshift, galaxies are more disturbed and gas less organised, but the global idea is that warm neutral medium can be widely spread, hence having a large total absorption cross-section. In turn, a mixture of warm and cold gas is found in more confined regions, closer to galaxies (where the pressure due to the total weight of material is higher) or in regions compressed by interactions. These are also the places where we expect molecular gas to be and subsequent star-formation to occur. We discuss this in the following sections.

3 Molecular gas

The abundance of molecular hydrogen depends on the equilibrium between formation and destruction of the molecule and has been studied intensively both theoretically and observationally in the local interstellar medium (see review by [Wakelam et al. 2017](#)). In short, H_2 can be formed by three-body reactions on the surface of dust grains (e.g. [Hollenbach & Salpeter 1971](#); [Jura 1974a](#)), but also in the gas phase through the $\text{H}^- + \text{H} \rightarrow \text{H}_2 + \text{e}^-$ reaction ([Black et al. 1987](#)). It happens that the gas-phase process is very inefficient and the production on the surface of dust grains actually dominates the formation rate even at low metallicity. The formation rate thus depends on the product of the hydrogen density to that of dust grains. Since at first approximation the abundance of dust follows that of metals (Z), one can see that the formation rate of H_2 is proportional to $n^2 Z$. This happens to be the same dependence as the cooling of the gas (Sect. 2.4.1), and already shows that H_2 is efficiently formed in the CNM only. H_2 is primarily through a two-step process, called Solomon process (see [Stecher & Williams 1967](#); [Dalgarno & Stephens 1970](#)): the absorption of UV photons in the electronic Lyman-Werner (LW) bands of H_2 is followed by a radiative decay onto the vibrational continuum (dissociative state) in approximately 12% of the cases ([Abgrall et al. 1992](#)). Because of this, self-shielding plays a crucial role and becomes very efficient when H_2 absorption lines from several rotational levels are saturated (e.g. [Draine & Bertoldi 1996](#)). Dust grains also absorb Lyman and Werner band photons (11.2-13.6 eV) and therefore contribute to decrease the photo-dissociation rate.

Theoretical works that include detailed treatment of the formation of H_2 onto dust grains and the dust- and self-shielding of H_2 show that a conversion from atomic to molecular hydrogen occurs above a $N(\text{H I})$ -threshold that increases with decreasing metallicity (e.g. [Krumholz et al. 2009](#); [McKee & Krumholz 2010](#); [Gnedin & Kravtsov 2011](#); [Sternberg et al. 2014](#)) and that the gas traced by H_2 is in a CNM phase. The basic picture is then that CNM clouds have a onion-like structure, with external, mostly atomic gas layers in which UV photons are progressively absorbed by the dust contained in the gas. Self-shielding of H_2 occurs quite abruptly, and the core of the clouds can be mostly molecular. Of course, if the cloud is very small and/or of low dust content, then the transition may not occur before reaching the cloud centre and the later has no molecular core at all.

Observationally, UV absorption spectroscopy of Galactic clouds towards nearby stars revealed that the molecular fraction, $f = 2\text{H}_2 / (2\text{H}_2 + \text{H I})$, sharply increases above a H I column density threshold of $5 \times 10^{20} \text{ cm}^{-2}$. A similar threshold has been found by [Reach et al. \(1994\)](#) from far-infrared emission studies of interstellar clouds, using dust as a tracer for H_2 . Higher column-density thresholds were observed in the Magellanic Clouds ([Tumlinson et al. 2002](#)), which could be the consequence of a higher UV radiation field together with a lower metallicity in these environments. However, it is also possible that a significant fraction of the observed H I column density is actually unrelated to the atomic envelopes of the H_2 -absorbing clouds ([Welty et al. 2012](#)), since $N(\text{H I})$ is derived through unresolved 21-cm emission, while $N(\text{H}_2)$ is measured in absorption.

Studying molecular hydrogen in different environments, and in particular at high-redshift is fundamental to understand the transformational processes of the gas that eventually lead to star formation, and the physical conditions in which these occurs. H_2 is therefore a key species for the understanding of galaxy formation and evolution. In the following, I summarise different efforts to search for H_2 in high-redshift DLAs (Sect. 3.1) and show how different selection techniques (Sect. 3.2 and 3.3) shed light on the relation between molecular gas and its environment. I also present a novel and promising technique to derive the column density distribution of H_2 from average statistical signal in low resolution spectra (Sect. 3.4). In Section 3.5, the peculiar relation between H_2 and neutral chlorine is discussed. I show how it is possible to derive the physical conditions in the molecular gas in Sect. 3.7 and how it is possible to use molecules to constrain fundamental physics and cosmology in Sect. 3.8.

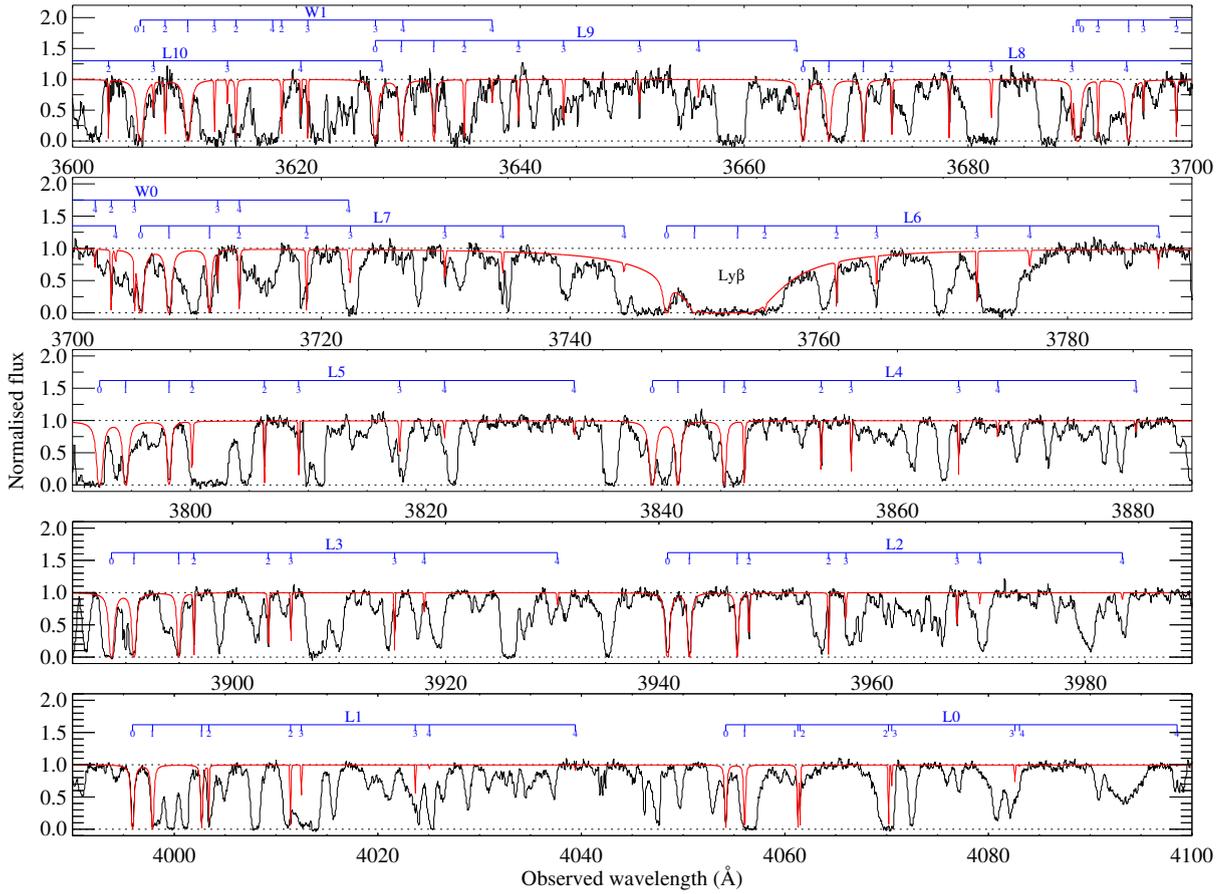


Figure 11: Portion of the high resolution VLT/UVES spectrum of the quasar J0643–5041 ($z_q = 3.09$) featuring intervening H_2 absorption lines at $z_{\text{abs}} = 2.6586$. Horizontal blue segments connect rotational levels (short tick marks) from a given Lyman (L) or Werner (W) band, as labelled above. The band number corresponds to the vibrational level of its upper level (ν'), while that of the lower level is here always $\nu = 0$.

3.1 Searches for H_2 in high-redshift DLAs

3.1.1 Using high spectral resolution

The systematic search for H_2 in damped Lyman- α systems was the main purpose of my PhD thesis. I briefly recall here the main results. In this work, we used the high-resolution spectrograph UVES on the Very Large Telescope to systematically search for H_2 in a large sample of DLAs (Noterdaeme et al. 2008b). The use of a blue sensitive and high-resolution spectrograph was determinant: the electronic transitions of H_2 have rest-frame wavelengths shorter than 1150 \AA , which also means that H_2 lines occur in the Lyman- α forest, see example in Fig. 11. The lines are generally noted as $\text{H}_2 \text{L}\nu' - \nu \text{B}J_l$ (for Lyman transitions) and $\text{H}_2 \text{W}\nu' - \nu \text{B}J_l$ (for Werner transitions), where ν' is the vibrational state of the upper level and ν that of the lower level. J_l is the lower level rotational state and B is the branch (P, Q or R) that corresponds to the selection rules between rotational levels $J_u - J - l = -1, 0, +1$.

The exceptional capabilities of UVES allowed us to obtain accurate column density measurements and very stringent upper-limits for H_2 non-detection, down to about $\log N(\text{H}_2) \sim 14$. Such excellent detection limits are not yet reached by any other search. Finally, the sample of 77 DLAs observed with UVES constituted the largest systematic search for H_2 and remains a reference for H_2 statistics in DLAs. The main results are the following: Firstly, we showed that H_2 is detected in about 10% of the DLA systems at $z > 1.8$. If we

restrict to the strong H₂ systems (in a self-shielded regime), with column densities $N(\text{H}_2) > 10^{18} \text{ cm}^{-2}$, then the detection rate is about $7 \pm 4\%$. This means that the total covering fraction of H₂ is small in DLAs and that the total cross-section of DLAs on the sky is likely dominated by the WNM phase.

Secondly, we found that the *overall* molecular fractions, $f = 2N(\text{H}_2)/(2N(\text{H}_2) + N(\text{H I}))$, are typically of less than 1% even in case of detection. This suggests that a large fraction of the atomic gas giving rise to the damped Lyman- α absorption is actually not related to the H₂-bearing gas.

Thirdly, we found that the H₂ detection rate is higher among DLAs with high metallicities (see also [Petitjean et al. 2006](#)) and high depletion factors, both being indicators of high dust abundance. We indeed further showed that H₂ is frequently detected when the column density of dust, estimated from that of iron missing from the gas-phase, is high. This clearly showed that dust is crucial for the formation and survival of H₂.

Finally, we found that the H₂ detection rate depends weakly on the H I column density, without any sharp change in behaviour, contrary to what is seen in our Galaxy. This can be explained by a large fraction of the gas not being related to the envelope of the H₂ cloud and by the fact that DLAs span a range of metallicities and physical conditions. This smooth out any sharp increase as a function of $N(\text{H I})$. We suggested that the typical column density above which the gas becomes predominantly molecular should have a higher value than that probed by our sample at that time.

3.1.2 Medium and low spectral resolution

Several years later, [Jorgenson et al. \(2014\)](#) performed a search of H₂ at $z > 2.2$ in a slightly larger sample (86 DLAs), all selected from the SDSS. They used different instruments, with a large fraction of observations performed at medium resolution ($R \sim 4000$) with the MagE spectrograph mounted on the Magellan telescope. From the single detection obtained (and already known from the UVES sample) these authors argued for a lower covering factor of strong H₂ systems in DLAs of 1% only (or $<6\%$ at 2σ). However, the tension with the results from the UVES sample is not as strong as actually presented by these authors. First of all, their detection limits vary over a wide range from $\log N(\text{H}_2) \sim 14-18.5$ at 1σ . Statistically speaking, the detection rates of both studies actually remain consistent with each other. In addition, the limits themselves obtained at $R \sim 4000$ remain questionable. Importantly, the selection function of both searches are also different (e.g. the redshift distributions differ by about $\Delta z \sim 0.3$) and biases can remain underestimated. For example, if the $N(\text{H I})$ -dependence of the H₂ detection rate is higher than what the UVES sample suggests, then the correction applied to mimic the true $N(\text{H I})$ distribution may not be enough. It is also possible that the quasar sample is biased against molecular-rich systems due to brightness requirements.

Interestingly, it is possible to detect H₂ absorption directly in SDSS spectra *provided the column density is high enough for the lines to be damped*. The first such detection was revealed accidentally by my DLA-searching algorithm: the spectrum of J0843+0221 appeared to have regularly-spaced damped absorption, which were actually the signature of single extremely strong H₂ absorber. We then developed a method to systematically search for similar damped H₂ systems in the SDSS-II spectra of DLAs ([Balashev et al. 2014](#)). The H₂ detection rate was found to be $\sim 9\%$ at $\log N(\text{H}_2) \sim 18.5$ ($\sim 7\%$ at $\log N(\text{H}_2) \sim 19$). However, we cautioned that the low quality of SDSS spectra leads the automatic procedure to overestimate $N(\text{H}_2)$, meaning that the derived values rather represent upper limits on the actual incidence rates for these column densities. Even after considering that the observed incidence rates actually apply to lower H₂ columns, this indicates an incidence rate higher than claimed by [Jorgenson et al.](#), and in better agreement with that derived from the UVES database.

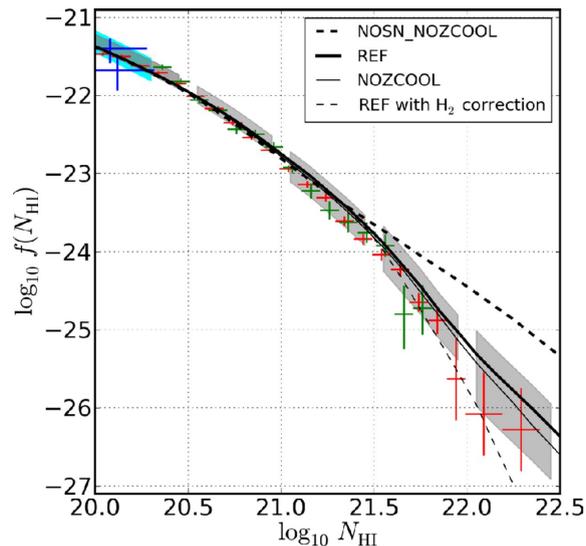


Figure 12: Column density frequency distribution in different simulations with or without feedback or H₂ compared to observational data from (O’Meara et al. 2007, blue), (Prochaska et al. 2010, cyan shaded regions), (Prochaska & Wolfe 2009, green) and (Noterdaeme et al. 2012b, red). Without H₂ or feedback, the distribution would follow a single power law (thick dotted). Models with stellar feedback or molecular hydrogen have a steeper slope at the high H I-column density end, in much better agreement with observations. This regime is subject to the most interesting physics and naturally calls for searching H₂ in high column density DLAs. [From Altay et al. 2013]

3.2 The high column density gas

Very high column density atomic gas should help in understanding the transition from atomic to molecular gas. Transition theories based on local micro-physics predict that the column density of atomic gas in the envelope of an H₂ cloud inversely scales with the gas metallicity, a dependence which is also seen in nearby galaxies (Schruba et al. 2018). Therefore, the typical H I column of H₂ clouds should be larger at high-*z* (where metallicities are lower, see Sect. 2.3.2) than it is locally.

There is also another reason to consider very high column density systems: since the gas falls in the gravitational potential of galaxies, we can naturally expect that higher gas column densities will trace more central regions of galaxies, closer to where stars are. This is reflected in the anti-correlation between the H I column density and the impact parameter (between the centroid of a galaxy and the point where the H I column is measured), as predicted by simulations (e.g. Pontzen et al. 2008) and observed over a wide range of redshifts (Zwaan et al. 2005; Rao et al. 2011; Krogager et al. 2012, see also Sect. 5). Simulations (e.g. Altay et al. 2013; Bird et al. 2014) also show that the high end of the $N(\text{H I})$ distribution is sensitive to stellar feedback and to the formation of molecules, as illustrated on Fig. 12.

Based on these considerations, I initiated a study of the systems with highest H I column densities ($\log N(\text{H I}) \sim 22$) in the SDSS. These systems are very rare, but the huge increase in the number of available quasar spectra made possible the detection of about a hundred of systems with $N(\text{H I}) \geq 5 \times 10^{21} \text{ cm}^{-2}$ (Noterdaeme et al. 2014). This is in an appreciable sample size when only a couple of them were previously known. Interestingly, several of the damped H₂ systems we detected directly in the SDSS (Balashev et al. 2014) are actually associated to these extreme DLAs, already suggesting a higher H₂ detection rate at high $N(\text{H I})$.

We have then obtained follow-up spectroscopic data with the VLT for a number of systems. In a Letter to the Editor (Noterdaeme et al. 2015a), we revisited the H₂ detection rate in DLAs as a function of H I column density, with statistics extended to the highest H I column density observed in intervening DLAs. In order not to bias the detection statistics, only systems selected upon H I and without any prior knowledge of H₂ content

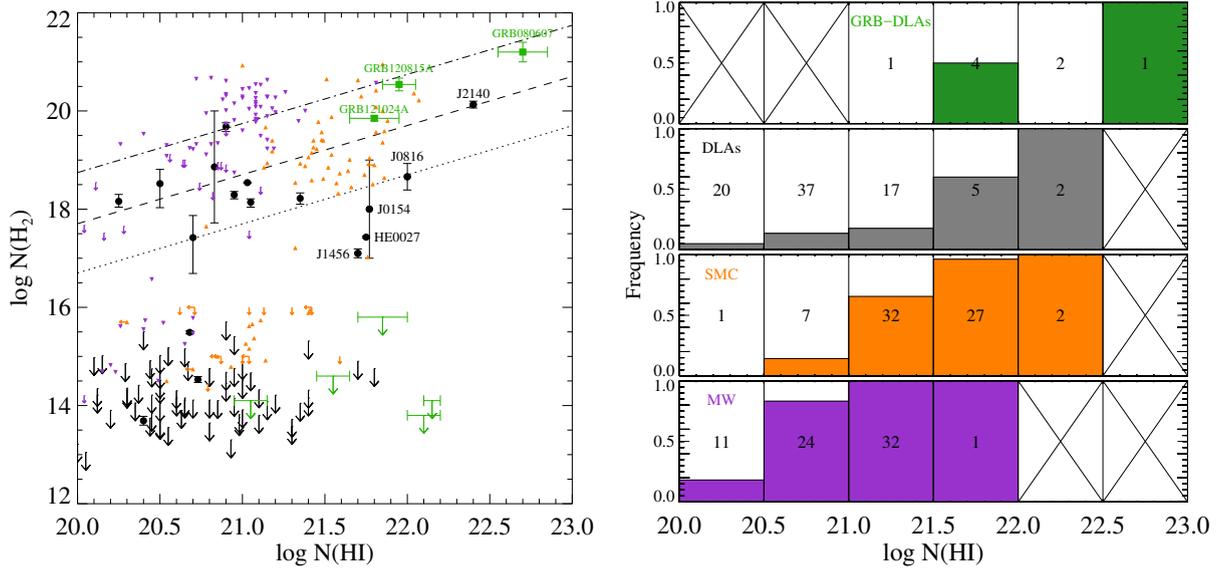


Figure 13: *Left:* Column density of H₂ as a function of that of H I. Black points and limits represent high redshift DLAs, Green points and limits correspond to DLAs associated with GRB afterglows. Purple (resp. orange) points represent measurements in the Milky Way (resp. SMC). The dotted, dashed, and dash-dotted lines represent average molecular fractions of, respectively, $f = 0.1$, 1, and 10%. *Right:* Frequency of H₂ detection as a function of the H I column density. For each $N(\text{H I})$ bin, systems with $\log N(\text{H}_2) \geq 17$ are coded with the same colour as in the left panel, and systems with $\log N(\text{H}_2) < 17$ are shown in white. Systems with upper limits that are not stringent (i.e. above $\log N(\text{H}_2) = 17$) are not taken into account. The numbers in each box indicate the total number of systems contributing to the bin. Crossed boxes have no statistics. [From [Noterdaeme et al. \(2015a\)](#)]

were considered. The main results, illustrated on Fig. 13, are that the detection rate of H₂ is now seen to increase with H I column density. While there is no abrupt transition, the detection rate appears to increase more steadily starting at roughly $\log N(\text{H I}) \sim 21.5$. This reflects both the spread in chemical and physical conditions in DLAs –as opposed to the more homogeneous metallicities seen in the Solar neighbourhood– and is consistent with the approximate inverse-metallicity scaling expected from the transition models. However, the overall molecular fractions remain very modest for most systems, even with high $N(\text{H I})$. This suggests that a significant fraction of the measured H I column density in this sample may still not be associated to the envelope of H₂ clouds. Taken as face-value, the higher H₂-detection rate at high $N(\text{H I})$ could rather be the consequence of the line of sight passing closer to the centres of the galaxy hosts, where the gas pressure is higher. We will discuss this further when deriving the physical conditions in the gas. We also found that the colour changes induced on the background quasar by continuum (dust) and line absorption (H I Lyman and H₂ Lyman & Werner bands) in DLAs with $N(\text{H I}) \sim 10^{22} \text{ cm}^{-2}$ and metallicities $\sim 1/10$ solar is significant, but not responsible for the long-discussed lack of such systems in optically selected samples. This is instead more likely due to small statistics. However, as we will discuss in Sect. 4.3, colour biasing can be severe at higher metal column densities and may bias the sample against the systems with high total H₂ content.

We have now obtained VLT/Xshooter spectroscopy for an extended and complete sample of extremely strong DLAs. This makes the basis for the project of my current PhD student, Adarsh Ranjan. We already singled out the detection of the highest column density of H₂ ever measured in an intervening DLA, with $\log N(\text{H}_2) = 21.3$ ([Ranjan et al. 2018](#)). The H₂ lines are indeed so strong that they completely dominate the absorption all over the Lyman- α forest. This single system also allowed us to test the H I-H₂ transition theories (see also Sect. 3.7) and discuss the link between the galaxy and the gas probed at very small galacto-centric radius (see Sect. 5.2.2). Analysis of the full sample is under progress and we expect exciting results.

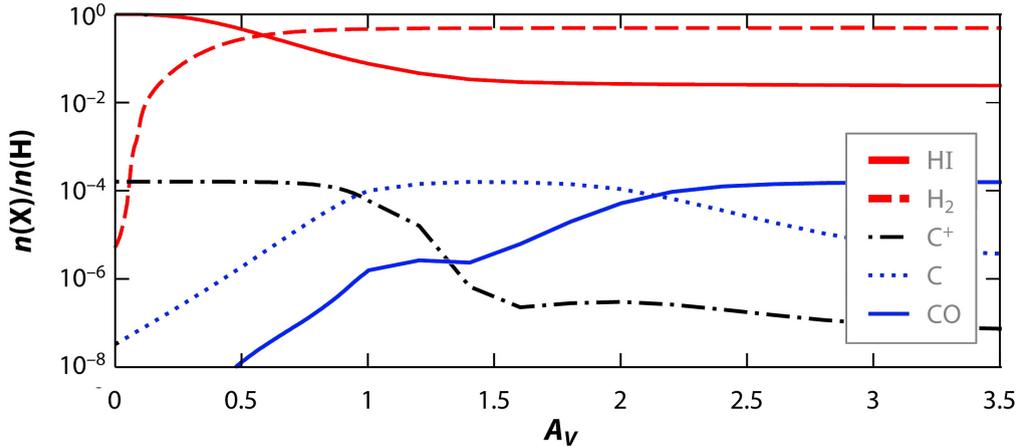


Figure 14: Structure of an interstellar cloud immersed into a UV radiation field. The axis-scales are indicative only. [Adapted from Bolatto et al. (2013)]

3.3 Flagging high-metallicity molecular gas using neutral carbon

Neutral carbon is known to be an excellent tracer of H_2 gas in our Galaxy (e.g. Snow & McCall 2006), in part because the ionisation energy of C I (11.3 eV) is close to that of Lyman-Werner photons that lead to H_2 dissociation⁶, but also because the carbon recombination rate increases with increasing density and so does the formation rate of H_2 . The relative abundances of hydrogen in atomic (H I) and molecular forms (H_2) along with that of carbon, in ionised (C II or C^+), neutral atomic (C I or C^0) and molecular (CO) forms are shown as a function of depth into the cloud in Fig. 14. The transition from H I to H_2 occurs quite abruptly. Carbon is found mostly in ionised in the external (atomic) layers of the cloud and continues as the dominant form of carbon even after the H I - H_2 transition. Although C I starts dominating over C II only deep in fully molecularised regions, its abundance increases as soon as H_2 starts to be formed and acts as a good tracer of the molecule. The abundance of CO also follows that of C I and ultimately dominates the overall carbon content in the very interiors of the cloud. Unfortunately, the corresponding phase is out of reach using UV absorption spectroscopy.

3.3.1 The survey for C I in the SDSS

Neutral carbon produces resonant absorption lines, out of which those at 1560 and 1656 Å are generally located out of the Ly- α forest, making it possible to search for strong C I absorption even at low spectral resolution. I searched for these lines at $z_{\text{abs}} > 1.5$ in SDSS spectra using a Spearman’s correlation analysis, after deriving the unabsorbed continuum using Savitsky-Golay filtering. In order to ascertain the detection, the presence of $\text{Si II } \lambda 1526$ and $\text{Al II } \lambda 1670$ absorption lines was also required. This should not introduce any bias since these metal lines are systematically detected in DLAs and much stronger than C I . An example of C I system found in the SDSS is shown on Fig. 15.

We then inspected the C I candidates by eye to remove spurious detections, and were left out with a sample of about 70 systems (Ledoux et al. 2015). We estimated our sample to be complete down to a rest-frame equivalent width limit of $W_{r,\text{lim}}^{\lambda 1560} \approx 0.4 \text{ \AA}$. We found that C I systems stronger than this limit are about 100 times rarer than DLAs at $z \sim 2.5$, but that their incidence rate (dN/dz) then increases significantly (by a factor 3) from $z \sim 2.5$ to $z \sim 1.5$ (see Fig. 16). Such an evolution is likely the combined effect of decreasing

⁶Carbon can still be ionised by UV photons with wavelengths in-between the H_2 lines.

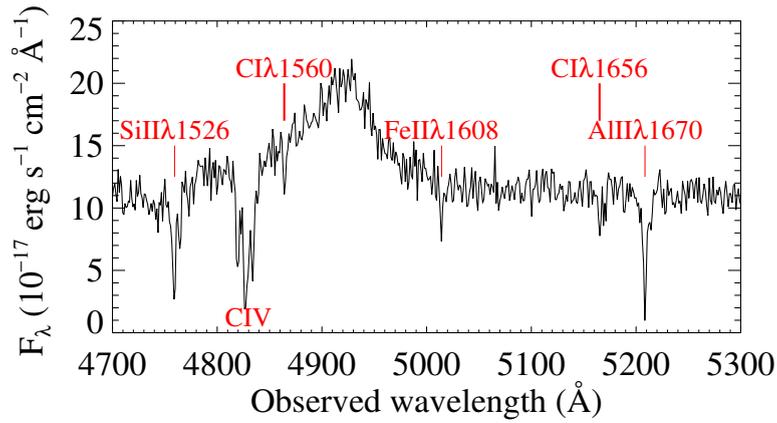


Figure 15: C I absorption lines detected at $z = 2.117$ in the SDSS spectrum of the quasar J1615+2648 ($z_{\text{em}} = 2.18$).

UV field (from both decreasing in-situ star-formation rates and decreasing UV background over this redshift range) and an increase of dust production in galaxies. Indeed, we found the C I systems to be systematically associated to a strong dust reddening of the background quasar. Since then, I have further developed the searching algorithm and applied the technique to an extended sample of quasar spectra from the SDSS-III and SDSS-IV database. To date, more than 150 strong C I systems are detected (unpublished).

3.3.2 Follow-up with the Very Large Telescope

We started following-up strong C I systems with the VLT as soon as we identified the first C I systems. Our pioneering high-resolution spectroscopic observations with UVES (under Director’s Discretionary Time) resulted in the first detection of CO molecules in absorption at high redshift (Srianand et al. 2008b), when this species remained elusive for 30 years of absorption studies. This motivated subsequent observations of C I systems (without any prior of the H I and H₂ column densities) which resulted in several additional CO detections (Noterdaeme et al. 2009b, 2010a, 2011, 2017, 2018; Ma et al. 2015). I will not detail here the properties of each system, but instead discuss the overall properties of strong C I systems, as derived from the complete spectroscopic survey with the VLT (Noterdaeme et al. 2018). We obtained spectra for almost all systems that are observable from Paranal Observatory, that is, a sample of thirty nine confirmed C I absorbers observed using UVES and/or X-shooter. X-Shooter was used because of its increased blue sensitivity, allowing observations of the faintest or lowest- z targets in our sample⁷. We showed that since CO lines are located out of the Ly- α forest, they can be detected with X-shooter with roughly the same detection limit as obtained with UVES, but using less telescope time. The drawback is that the kinematic structure and the rotational levels cannot be resolved at the medium spectral resolution achievable with X-shooter.

The main result of the survey are shown in Fig. 17: H₂ is detected in all twelve systems where the corresponding lines are covered by the instruments setup, with $N(\text{H}_2) \geq 10^{18} \text{ cm}^{-2}$, indicating a self-shielding regime. However, the H₂ column density does not strongly correlate with that of C I (middle panel) while the overall molecular fractions tend to increase more clearly with $N(\text{C I})$ (bottom panel). In addition, the molecular fractions are significantly higher than other strong H₂ systems from the literature, that were selected classically as DLAs (i.e. H I-selected). All this can be understood in the context of H I-H₂ transition theories together with selection effects. Indeed, Siwei Zou, who just defended her PhD under supervision by Patrick Petitjean and

⁷Other advantages of X-shooter is that it allows us to simultaneously derive the dust-reddening over the full-wavelength range from UV to NIR and search for species with absorption lines in the NIR (Zou et al. 2018) as well as looking for nebular emission lines from the associated galaxies (see Sect. 5).

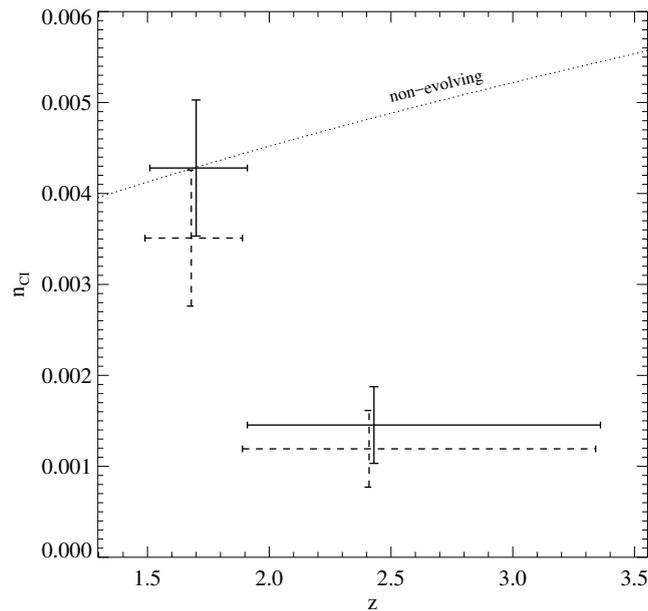


Figure 16: Evolution of the incidence of intervening strong C I absorption systems. The measurements (solid) are corrected for incompleteness at our detection limit of $W_r(\text{C I}1650) = 0.4 \text{ \AA}$, with the uncorrected data shown as dashed error bars. The curve represents the expected incidence for a non-evolving population arbitrarily scaled to the value at $z = 1.7$. [From Ledoux et al. 2015]

myself, found that C I systems have metallicities close to Solar, i.e. much higher than the overall DLA population (Zou et al. 2018). This means that their critical H I column density (beyond which any extra hydrogen added to the cloud is converted to H₂) is less than that of lower-metallicity molecular systems. In addition, C I systems were selected without any prior on the H I column meaning that unrelated H I gas (from e.g. the diffuse warm gas) will be statistically less present than in the H I-selected systems (see indeed Sect. 3.2).

CO is detected in 18% (7/39) of the systems down to a detection limit of $\log N(\text{CO}) \sim 13.5$. We found $N(\text{CO})$ and $N(\text{C I})$ to be strongly correlated, with a ratio of one to ten. Since the detection limits are significantly below (1 dex) the CO column densities measured in case of detection, this result is robust. This suggests that the C I-selected population is probing gas deeper than the H I-H₂ transition, and means that the gas probed by C I has high *local* molecular fractions. The strong correlation between CO and C I is expected since their respective abundances both depend on the dust-shielding and on the abundance of metals. Interestingly, the CO/C I ratio ($\sim 1/10$) is significantly higher than seen in most models of translucent clouds (which predict a ratio of about 1/100, see e.g. Fig. 14), an issue already raised by Sonnentrucker et al. (2007). Several factors such as dust grain size distribution and physical conditions in the cloud (including cosmic rays) likely play important roles in setting these factors (see e.g. Shaw et al. 2016 and Noterdaeme et al. 2017). Detailed studies of the chemical and physical conditions in the molecular gas are crucial to understand this better (see Sect. 3.7).

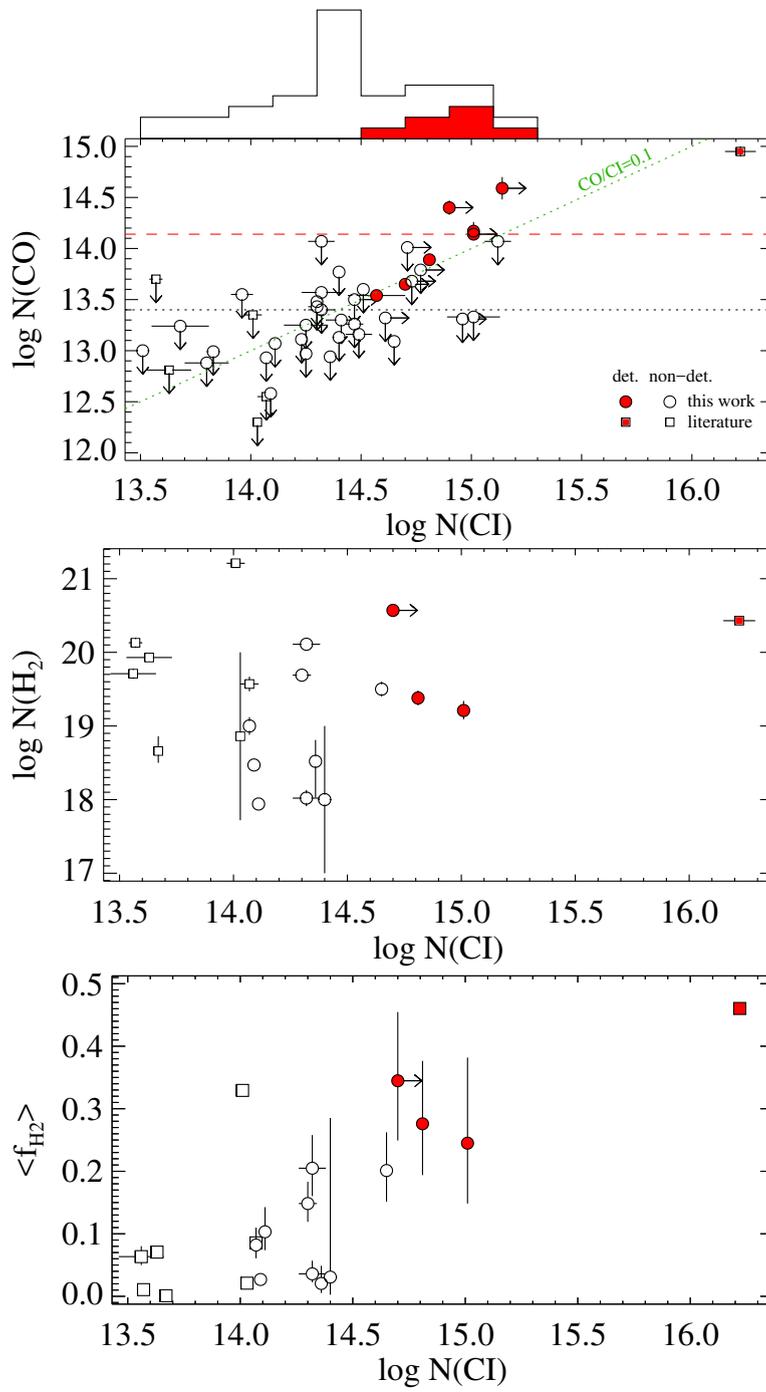


Figure 17: Observed column densities of CO (top), H₂ (middle) and overall molecular fractions (bottom) as a function of the column density of neutral carbon. CO detections are represented by red colours. The $N(\text{C I})$ -distributions and median $\log N(\text{CO})$ values (horizontal lines) are shown for the statistical sample only (circles). Squares correspond to high- z H₂ DLA systems from the literature (Albornoz Vázquez et al. 2014; Balashev et al. 2010, 2011, 2015, 2017; Carswell et al. 2011; Guimarães et al. 2012; Klimentko et al. 2016; Noterdaeme et al. 2007b, 2008b, 2015b, 2017; Petitjean et al. 2002). [From Noterdaeme et al. 2018]

3.4 The H₂ distribution function from average H₂ signal in SDSS

While the detection of H₂ is eased by high resolution spectroscopy with $R \sim 50\,000$ (e.g. [Levshakov & Varshalovich 1985](#); [Ledoux et al. 2006b](#)), this technique remains extremely time-consuming for the overall population of DLAs because of the low detection rate ($\lesssim 10\%$). In other words, a large number of DLAs needs to be surveyed to constrain H₂ statistics. The largest such sample (Sect. 3.1.1) was built using archival plus dedicated data but selection effects are hard to control, in particular because of the limited statistics when splitting into sub-samples. Blind follow-up of DLAs using medium-resolution spectroscopy ([Jorgenson et al. 2014](#)) seems not to be a good strategy since it still remains costly in terms of observing time on large telescopes. Furthermore, the improvement in terms of resolution reached by these authors is only a factor of two compared to the SDSS data, hampering the ability to detect H₂ at levels comparable to high- R data (UVES has a factor of 10-15 higher resolution). A selection based on metallicity ([Petitjean et al. 2006](#)) allows to increase the H₂ detection rate but this is not of very practical use since metallicities are hard to estimate from low-resolution data. Our subsequent selection based on the sole presence of C I lines in the low resolution spectra is in turn very efficient (100% detection rate). However, it selects the population of metal-rich systems only, so it does not provide the global H₂ incidence in the overall neutral gas. Finally, directly searching for H₂ at the DLA redshifts in individual low-resolution spectra allows one to directly take advantage of large database without the need of follow-up, but this method only allows to detect very strong systems and the column-density measurements remain difficult because of blending with the Lyman- α forest unless follow-up observations are also obtained.

For these reasons, we have very recently developed a new technique based on stacked data. The idea is that one can obtain directly information on the mean properties of a population of quasar absorbers by analysing composite spectra built by averaging a large amount of individual spectra with low S/N and intermediate resolution (e.g., [Nestor et al. 2003](#); [Wild et al. 2006](#); [Noterdaeme et al. 2010b](#); [Rahmani et al. 2010](#); [Joshi et al. 2017a, 2018](#); [Mas-Ribas et al. 2017](#)). A main strength of this method is that a feature that cannot be confidently detected in individual spectra or that is present only in a fraction of them becomes visible in the high S/N composite spectrum. Here, we used stacked SDSS spectra of DLAs to detect the weak average H₂ signal. One particularity of the H₂ is that its formation-dissociation equilibrium depends strongly on the self-shielding, which changes steeply around $\log N(\text{H}_2) \sim 16$, when the lines get saturated. At this column density, a slight increase of the column density of H₂ strongly increases the shielding, itself leading to a higher column density. Conversely, a small decrease in the H₂ column will decrease significantly the shielding, leading to a rapidly increasing dissociation rate. Observationally, the distribution of H₂ column densities in different environments (e.g. Solar neighbourhood, Magellanic clouds and high- z DLAs) is indeed found to be bimodal, with values either above $\log N(\text{H}_2) \sim 18$ (self-shielded regime) or well below $\log N(\text{H}_2) \sim 14$. This means that, for a given system, the H₂ signal in SDSS will be either very weak, hence considered as non-detection, or strong enough to produce intrinsically saturated lines. It is then possible to derive the H₂ covering fraction from the mean H₂ profile in the high-S/N stacked spectrum, see Fig. 18.

Our results, presented in a Letter to the Editor ([Balashev & Noterdaeme 2018](#)), are the following: we derive a covering factor of strong H₂ ($\log N(\text{H}_2) > 10^{18} \text{ cm}^{-2}$) to be about 4% in systems with $N(\text{H I}) > 10^{20} \text{ cm}^{-2}$, about 9% in a "metal" sub-sample with prominent metal lines and about 30% in systems with $N(\text{H I}) > 5 \times 10^{21} \text{ cm}^{-2}$. The total DLA covering factor is consistent with both the results from the UVES ($7 \pm 4\%$) and MagE ($< 6\%$) searches. It could however be slightly underestimated compared to bona-fide DLAs since the stack also include strong sub-DLAs (with $20 \leq \log N(\text{H I}) < 20.3$). The higher detection rate in the metal sample is in agreement with our previous observation that H₂ is more frequently detected at high-metallicity. The strong increase of the H₂ incidence rate at high H I column densities is also consistent with steady-state models for H I-H₂ conversion.

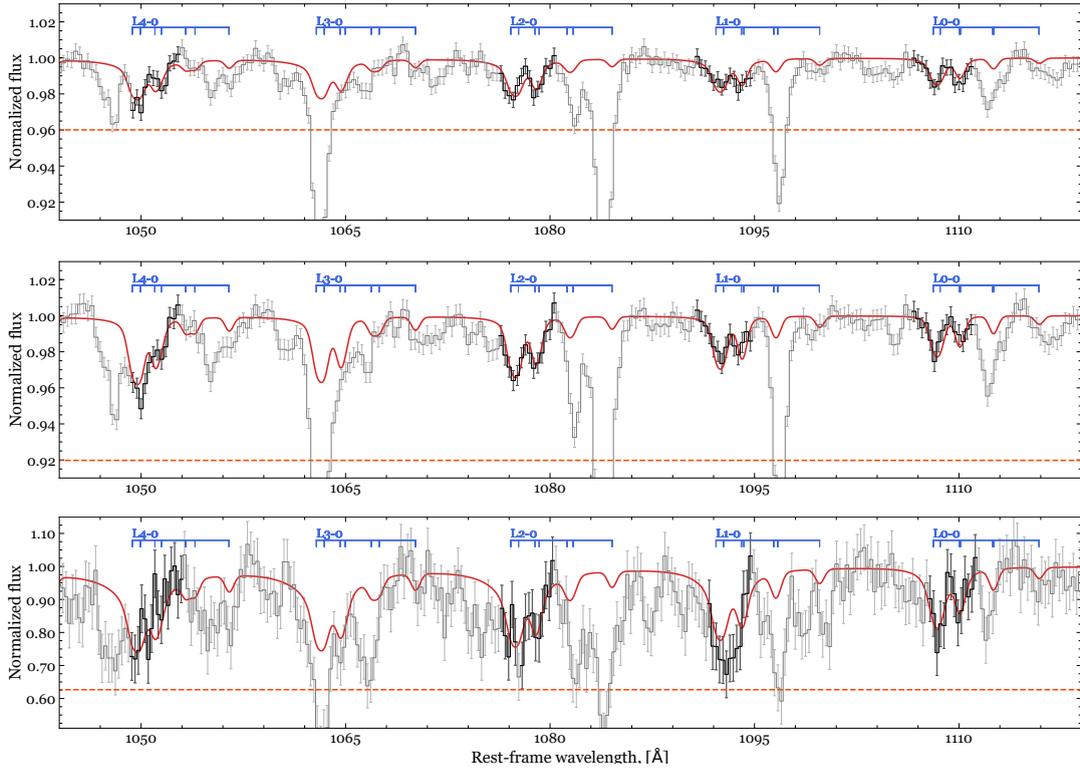


Figure 18: Statistical detection of H_2 in three composite DLA spectra, built from SDSS-III/BOSS. From top to bottom: *all DLAs*, *metal and ESDLAs* samples. Note the much wider y-axis range for the *ESDLA* composite. *Left panels:* black lines represent the H_2 $J=0,1$ regions that were used to constrain the fit (i.e. not blended with other absorption lines such as $Fe\ II$ at 1063, 1081, 1096 Å), the rest of the composite spectra being shown in grey. Red lines show the best-fit composite H_2 profiles. The orange dashed horizontal lines correspond to the best-fit value of effective non coverage factor, $1 - r$, where r is the incidence rate of strong H_2 systems. [From Balashev & Noterdaeme 2018]

From the average signal, we derived the slope of the truncated power-law distribution used to model the H_2 distribution. We then obtained the H_2 frequency distribution in the range $\log N(H_2) \sim 18 - 22$, shown in Fig. 19. This figure also shows the $f_{H_2}(N)$ derived at $z = 0$ by Zwaan & Prochaska (2006) for $\log N(H_2) > 22$ using CO emission maps of nearby galaxies. Interestingly, the two distribution connect well to each other at $\log N(H_2) \sim 21 - 22$. This could indicate a self-regulating mechanism or be only coincidental due to several factors acting in opposite ways: On one hand the fraction of H_2 among per $H\ I$ -selected system is expected to be lower at high-redshift because of the higher UV background, the typically lower metallicities and the statistically lower pressures compared to well-formed local galactic discs. On the other hand, the total incidence of $H\ I$ decreases with cosmic time (i.e. DLAs are less numerous at low- z), so that the incidence of H_2 (which is the product of the incidence of DLAs times the H_2 detection rate in DLAs) remains almost constant, at least in that column density range. Speculating a bit further, by integrating $f_{H_2}(N)$ and extrapolating over higher column densities, we find that about 15% of the H_2 resides in diffuse molecular clouds (with $\log N(H_2) < 22$), i.e. that most of the H_2 in the Universe resides in very high column density systems, that still escape direct detection through UV lines.

From $f_{H_2}(N)$, we can also derive the atomic-to-molecular transition column density, which corresponds to the column density at which $f_{H_2}(N)$ equates $f_{H\ I}(N)$, here $\log N \sim 22.3$. This is of course a typical value for the average physical and chemical conditions in DLAs. This is consistent with the expected critical value for conversion into H_2 of $\Sigma_{H\ I}^{crit} \sim 10/Z M_\odot pc^{-2}$ (McKee & Krumholz 2010), where $Z \sim 1/20$ is the average metallicity of DLAs at $z \sim 3$ (see Sect. 2.3.2). Of course, individual systems may have transitions at very different values, depending on their chemical and physical conditions (for example, the transition occurs at

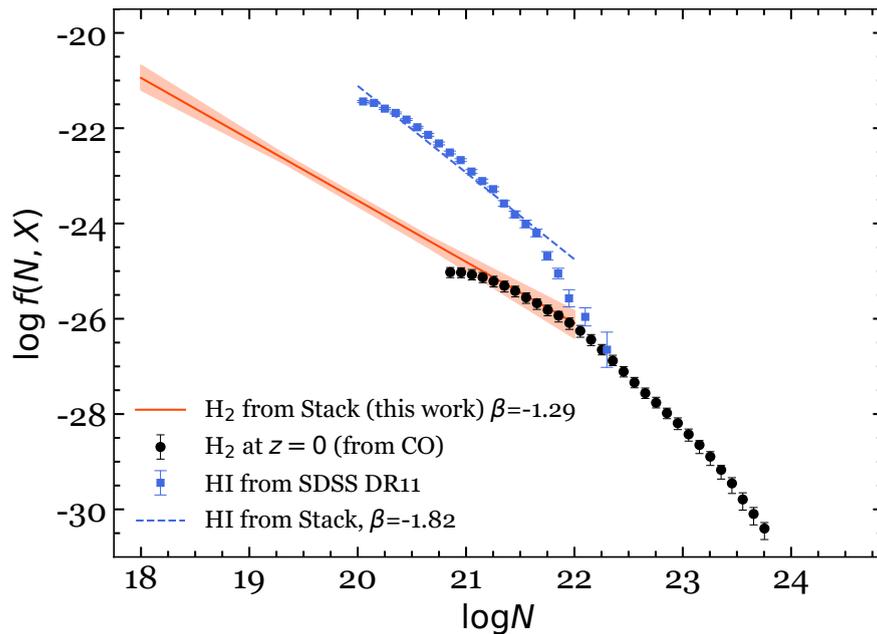


Figure 19: Column density distribution functions of H_2 and HI . The blue points correspond to the high- z HI distribution function from DR9 (Noterdaeme et al. 2012b) with the high $N(\text{HI})$ -end updated using DR11 (Noterdaeme et al. 2014). The blue dashed line is our derivation using the average Lyman- α absorption from the SDSS DLA composite (Mas-Ribas et al. 2017). The black dots correspond to the H_2 distribution function derived at $z = 0$ from CO maps Zwaan & Prochaska (2006). Finally, the red line with shaded uncertainties correspond to our measurement of the H_2 distribution function at high- z , also from the composite DLA spectrum. [From Balashev & Noterdaeme 2018]

$\log N \sim 20.8$ in the solar-metallicity system towards J0000+0048, Noterdaeme et al. 2017). At least, this result provides a first observational test to broadly confirm the transition theories in their simplest form. This is also useful to validate the recipes for H_2 formation used in hydrodynamical simulations of galaxy formation and evolution (e.g. Lagos et al. 2011).

We finally remark that while stacking spectra appears to be a promising way to gather information on the statistical properties of H_2 absorbers, this first work should be mostly considered as a proof of concept and a careful assessment of the biases and selection function is still required to make strong conclusions.

3.5 Chlorine chemistry

Among all metals found in the neutral gas, chlorine is a peculiar species. Its first ionisation potential (12.97 eV) is less than that of hydrogen, which implies that chlorine is easily ionised in neutral gas. Like C I, shielding by dust and H_2 contribute to decrease the photo-ionisation rate of chlorine, so that its presence in molecular clouds is not surprising. However, unlike carbon, chlorine is found almost exclusively in neutral form (Cl I) as soon as some H_2 is present. This is due to rapid chemical reactions. Indeed, chlorine reacts exothermically with H_2 at a very high rate, rapidly converting Cl^+ into HCl^+ , which subsequently releases neutral chlorine through several channels (Jura 1974b; Neufeld & Wolfire 2009). Neutral chlorine is therefore a good tracer of H_2 , even in diffuse molecular gas. This linking is actually very useful to constrain the velocity profile of strong (saturated) H_2 systems where individual components are hard to distinguish from H_2 lines alone. Moreover, chlorine is expected to deplete little onto dust grains so that the relative abundance of chlorine to that of hydrogen should be a good indicator of the metallicity in the H_2 -bearing gas. Conversely, if a measurement of the metallicity can be obtained independently (from e.g. the abundance of zinc), then it is

possible to constraint the local molecular fraction. The metallicity in the H₂-bearing gas can then be written as

$$[\text{Cl}/\text{H}] = [\text{Cl I}/\text{H}_2] + \log f_{\text{H}_2} \quad (10)$$

where

$$[\text{Cl I}/\text{H}_2] = \log \left(\frac{N(\text{Cl I})}{2N(\text{H}_2)} \right) - \log \left(\frac{\text{Cl}}{\text{H}} \right)_{\odot}, \quad (11)$$

and f_{H_2} is the local molecular fraction in the cloud. We studied the abundance of chlorine in strong H₂-bearing systems from the literature (mostly from our searches) and found that $N(\text{Cl I})$ correlates strongly with $N(\text{H}_2)$ (at 5σ confidence level) with only 0.2 dex dispersion over a wide range of H₂ column densities ($\log N(\text{H}_2) \sim 18 - 20$), i.e., down to 10 times lower column densities than those reached through the observations of nearby stars (Balashev et al. 2015). Surprisingly, the observed Cl I/H₂ ratios at high redshift are very similar to that observed in the Solar neighbourhood (Moomey et al. 2012). This can be the consequence of similarly high chemical enrichment and high molecular fractions in the H₂-bearing gas at high redshift or that both the local molecular fraction and the metallicity are linearly correlated.

We further found that $[\text{Cl I}/\text{H}_2]$ is systematically higher than the overall measured metallicity. This can be explained by a local molecular fraction less than unity. Assuming the metals are well mixed and the metallicity is constant throughout the system, this still implies that the local molecular fractions are typically an order of magnitude higher than the overall value. In other words, only one tenth of the H I is found in the diffuse molecular gas. Interestingly, those systems where $[\text{Cl I}/\text{H}_2]$ are closer to the observed overall metallicity are those where CO is also detected. The H₂ component is thus likely almost fully molecularised in these cases.

3.6 Deuterated molecular hydrogen (HD)

Like molecular hydrogen H₂, the deuterated isotopologue HD can be present in the interstellar medium. HD also produces a set of Lyman and Werner lines but its much lower abundance makes the corresponding absorption lines much weaker than that of H₂. In fact, the first high- z detection of HD was obtained only in 2001 by Varshalovich et al. at $z_{\text{abs}} = 2.34$ towards PKS 1232+082. About a dozen other HD systems have now been identified (Noterdaeme et al. 2008a; Ivanchik et al. 2010; Tumlinson et al. 2010; Balashev et al. 2010, 2017; Noterdaeme et al. 2017) and we expect this number to increase significantly with our ongoing surveys. If we call $f_{\text{HD}} = \text{HD}/(\text{HD} + \text{D I})$ the fraction of deuterium which is in molecular form, then we can write

$$\frac{\text{HD}}{2\text{H}_2} = \left(\frac{\text{D}}{\text{H}} \right) \frac{f_{\text{HD}}}{f_{\text{H}_2}} \quad (12)$$

where D/H is the intrinsic total abundance of deuterium.

If both H₂ and HD are predominantly in molecular form (i.e. $f_{\text{H}_2}, f_{\text{HD}} \approx 1$), then the HD/2H₂ ratio can provide an alternative to (D I/H I) measurements to constrain the primordial (D/H) ratio (Ivanchik et al. 2010), provided the gas metallicity remains sufficiently low for deuterium astration not to be significant yet (see Sect. 2.3.3). This is however generally not the case for H₂ absorption systems, although we have found cases with strong H₂ at low metallicity (Ivanchik et al. 2010; Balashev et al. 2017). Alternatively, the (D/H) ratio can be used to test chemical evolution models (e.g. Dvorkin et al. 2016). A main observational advantage of HD over D I is that the identification of the molecular absorption lines is unambiguous and the measurement of the column density is generally more accurate.

If one wants to use HD to constrain either $(D/H)_p$ or chemical evolution models, the main question is then to know whether f_{HD} equates f_{H_2} or not. Since the dissociation of both molecules occurs mainly through the Solomon process, we could expect f_{H_2} to be much higher than f_{HD} simply because H_2 self-shields much before HD. However, it happens that chemical fractionation can actually favour the opposite situation. Without entering into details of the different processes affecting the molecular deuterium-to-hydrogen ratio, which are reviewed by Liszt (2015), we note that the reaction $D^+ + H_2 \rightarrow HD + H^+$ is fast and can lead to an increase of HD compared to H_2 . In the case of the molecular system at $z_{abs} = 2.426$ towards SDSS J0000+0048 (Noterdaeme et al. 2017), we found $f_{HD}/f_{H_2} \sim 3$, assuming an intrinsic primordial (D/H) value (the ratio is even higher if significant astration is at play). This naturally implies that the cloud cannot be fully molecular (since $f_{HD} < 1$), a conclusion we also reached from deriving the physical conditions through numerical modelling.

There is growing interest in the community for studying HD in the local ISM since it provides constraints on the physical processes at play (the formation rate of HD, for example, depends on the density of protons/deuterons and hence on the strength of the radiation that can ionise hydrogen in the clouds interior, such as cosmic rays). Thanks to our optimised selection techniques, we are rapidly increasing the number of HD detections at high- z (not all published) where we can lead similar studies.

3.7 Physical conditions

As for the atomic gas (Sect. 2.4), the excitation of different species allow to derive the local physical conditions in the H_2 -bearing gas as well. Furthermore, in molecular gas, not only the survival of neutral and molecular species already provides constraints on the physical conditions, but these species also have a large number of excited states with different sensitivities on the density, temperature, UV flux, etc. We mainly discuss here the excitation of two species of particular interest: H_2 and C I.

3.7.1 Rotational excitation of H_2

Excitation diagrams are a common representation of the population of H_2 molecules across different levels. We illustrate that obtained for H_2 in the high- z DLA towards J2140–0321 in Fig. 20, but almost all observations of H_2 in absorption lead to similar diagrams (at least qualitatively), both in the distant Universe like in our own Galaxy. In local thermal equilibrium, the excitation of a given H_2 rotational level J compared to the ground state ($J = 0$) is described by

$$\frac{N(J)}{N(0)} = \frac{g(J)}{g(0)} e^{-E_{0J}/k_B T_{0J}} \quad (13)$$

where T_{0J} is the excitation temperature and $g(J) = (2J + 1)(2I + 1)$ is the quantum weight of level J , with I being the total nuclear spin and equal to 0 for para- H_2 (even rotational levels J , where the nuclear spins of the two hydrogen atoms are aligned) and 1 for ortho- H_2 (odd J , when the nuclear spins are in opposite directions). Radiative transitions between these two forms are highly improbable, but a conversion from ortho- H_2 to para- H_2 (or vice-versa) can occur through collisions (Dalgarno et al. 1973; Flower & Watt 1984; Mandy & Martin 1993; Le Bourlot et al. 1999) and possibly on the surface of dust grains (Le Bourlot 2000). It happens that the typical collisional time-scale in cold gas is located in-between the radiative lifetime of the low and the high rotational levels.

The low rotational levels have long radiative lifetime so that their population ratio is maintained by collisions

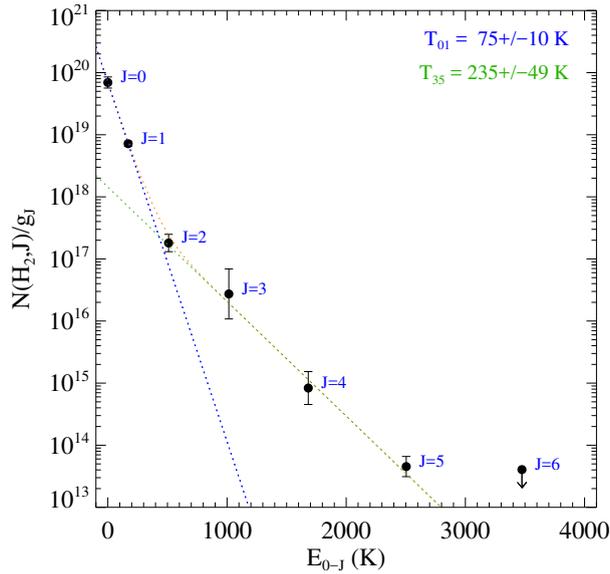


Figure 20: H_2 excitation diagram at $z_{\text{abs}} = 2.34$ towards SDSS J2140–0321. [From [Noterdaeme et al. 2015b](#)].

at the Boltzmann value. In fact, T_{01} excitation temperature has long been considered as an excellent measure of the kinetic temperature of the gas (i.e. $T_k \approx T_{01}$), at least when these levels are self-shielded. In the local ISM, this has been corroborated by comparing the H_2 -based temperatures with those obtained from 21-cm observations ([Roy et al. 2006](#)). The average temperature of H_2 -bearing clouds is found to be ~ 77 K in the Milky-Way ([Savage et al. 1977](#)). At high redshift, temperatures measured from the excitation of H_2 tend to be slightly higher, of the order of 100-150 K ([Srianand et al. 2005](#)), which could be due to lesser shielding from UV photons. We will discuss this further below.

High-rotational levels are not thermalised, as evidenced by the break in the excitation diagram. Two main mechanisms are known to excite these high J -levels: the so-called radiative and formation pumping. In the first case, after absorption of a UV photon, H_2 decays through fluorescence back into the electronic ground-state passing through different rotational and vibrational states. The population of H_2 is then redistributed into different rotational levels. In the case of formation pumping, H_2 is released from dust grains in a high rotational level after its formation. However, considering that the formation rate is balanced by the photodissociation rate, which corresponds to only 10-15% of UV photons absorbed by H_2 , one can see that the UV pumping (i.e. the 85% remaining) will dominate over the formation pumping by approximately a factor of five. The population of high rotational levels of H_2 can therefore in principle give constraints on the UV field ([Noterdaeme et al. 2007b](#)). The observation that the Doppler parameters of H_2 lines increases with their rotational level provides some additional hints about the excitation processes ([Noterdaeme et al. 2007a](#)). This indicates that high- J levels arise mostly from external layers of the cloud, where the UV field is less attenuated and the temperature can be higher, but turbulent dissipation or C-shocks could also play a role. Modelling the excitation of high- J levels is not an easy task. Alternatively, it remains instructive to compare the excitation of H_2 with that obtained in other environments where the physical conditions (in particular UV field) is known better. For example, H_2 is observed with a high excitation temperature of high- J levels close to the bright O-star HD 34078 ([Boissé et al. 2005](#)) when its excitation is lower on average in the Milky-Way disc ([Gry et al. 2002](#)).

3.7.2 Excitation of C I fine-structure levels

Neutral carbon's ground state is a triplet ($2s^2 2p^2 \ ^3P^e$) with the fine-structure levels, shortly denoted C I ($J = 0$), C I* ($J = 1$) and C I** ($J = 2$). C I fine-structure levels can be excited by collisions (mostly H^0 and H_2), UV pumping and direct excitation by CMB photons. At low densities, the excitation of C I depends strongly on the UV field, but this is not the case anymore for the typical densities of the cold gas (Silva & Viegas 2002), where C I excitation is dominated by collisions and CMB photons. Knowing the temperature of the CMB, this means that by measuring the population repartition of C I fine-structure levels, it is possible to determine the gas pressure. Absorption lines from the three levels are generally seen together and, because the corresponding lines are easily resolved (separation from a few to several tens km s^{-1}), the column densities in the different excited states can generally be measured quite accurately.

Since C I and H_2 have a high degree of co-spatiality, their excitation can be used together to constrain the temperature and density of the gas (Fig. 21.). Other species can then be used as independent check or simply to narrow down the constraints. From our studies, we found that in most cases the H_2 /C I-bearing gas has temperatures of the order of 100 K and densities of the order of 100 cm^{-3} , i.e. these are about the canonical values for the CNM in our galaxy.

3.7.3 Examples: H_2 at high and low metallicity

Interestingly, quasar observations allow us to probe broader ranges of chemical enrichment than in our Galaxy, which we illustrate by two cases: The H_2 absorber towards J0000+0048 (Noterdaeme et al. 2017), with super-Solar metallicity and high overall molecular fraction and the high column density H_2 system towards J0843+0221 (Balashev et al. 2017), that has only $Z \sim 0.03 Z_\odot$. The gas column density in the low-metallicity case is about ten times that in the high-metallicity one but this is not sufficient to reach same total dust shielding (which is, at first order, proportional to $N(H)Z$). Indeed, we found strong reddening towards J0000+0048 but little towards J0843+0221. The presence of molecular gas at such different metallicities illustrates the role also played by the volumic density. The formation rate of H_2 on dust grains depends on the densities of both hydrogen and dust grains, i.e. it is proportional to $n \times n_{\text{grains}} \sim n^2 Z$. Density influences then quadratically the H_2 formation rate. We indeed found $n \sim 300 \text{ cm}^{-3}$ towards J0843+0221 while H_2 exists with only $n \sim 80 \text{ cm}^{-3}$ towards J0000+0048, so that the H_2 formation rates become actually comparable. The high density in the low-metallicity H_2 cloud likely results from a high external pressure (this is also what we observe) and this is most likely to be found closer to the galactic centre, where a lot of gas has accumulated. This then explains the very high H I column density observed, in agreement with our proposition based on the detection of star-formation activity at low impact parameter, discussed in Sect. 5.2.1. In summary, the high column density helps here to decrease the dissociation rate of H_2 (although not very efficiently due to the low dust content), but more importantly, it probes high pressure regions, where the formation rate of H_2 is boosted even at low metallicity. At the same time, the thermal equilibrium is expected to be shifted to higher temperatures due to the low cooling rate through metal lines (see also Glover & Clark 2014). This is indeed what we observe, with T_{01} begin more than twice higher than what is observed in our Galaxy for the same column density. We finally note that the systems correspond to the so-called CO-dark molecular gas (see Wolfire et al. 2010). In the high metallicity case, CO absorption is detected, but the density and surface density is too low for significant emission. The surface density is similar to that of nearby ^{13}CO -selected clouds in the second case, but the low-metallicity impedes the presence of CO. In the local Universe, it has been shown that such molecular phases, while not being detectable through CO-emission, may contain a significant amount of the total gas mass in galaxies.

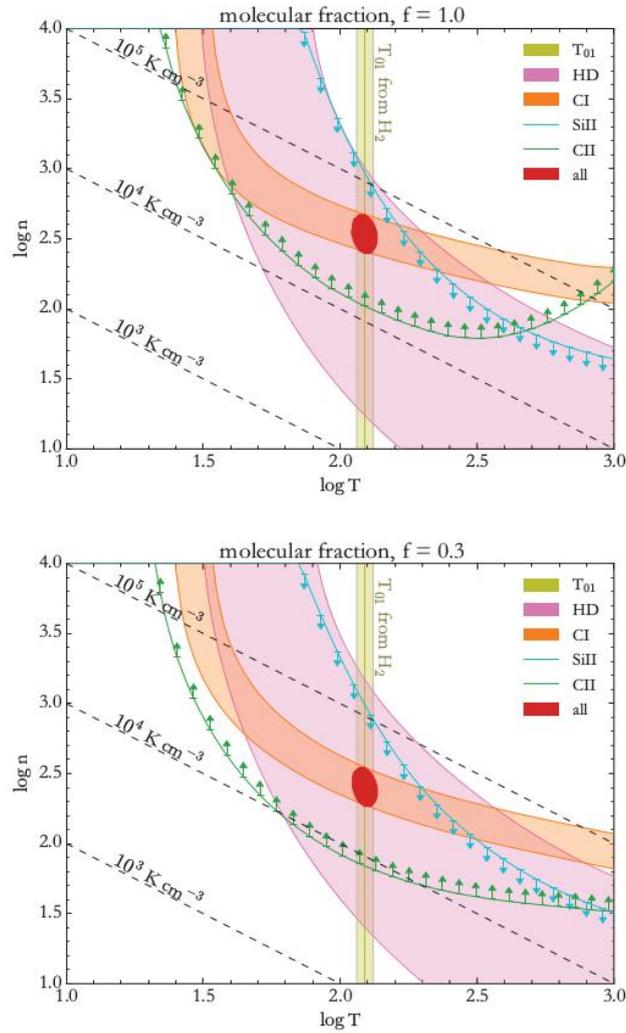


Figure 21: Constraints on the density-temperature parameter space from different species found associated to the molecular gas towards J0843+0221. The top and bottom panels assume two extreme cases for the molecular fraction, using the overall observed value throughout the system $f_{\text{H}_2} = 0.3$ (bottom panel), and considering full molecularisation (top panel). The actual local molecular fraction must be in-between these values. [From Balashev et al. 2017].

3.7.4 Analytical expression for the H I surface density of a H₂ cloud

Finally, theoretical works considering the microphysics of H₂ formation and destruction (Sternberg et al. 2014) have linked the the total H I column density shielding a H₂ cloud to the dust cross-section, the incident UV flux and the volumic density (see also Bialy et al. 2017) as:

$$\Sigma_{\text{HI}} = \frac{6.71}{\tilde{\sigma}_g} \ln\left(\frac{\alpha G}{3.2} + 1\right) M_{\odot} \text{pc}^{-2} \quad (14)$$

where $\tilde{\sigma}_g \equiv \sigma_g / (1.9 \times 10^{-21} \text{ cm}^2)$ is the dust grain Lyman-Werner (LW = 11.2 - 13.6 eV, 911.6 Å - 1107 Å) photon absorption cross section per hydrogen nucleon normalised to the fiducial Galactic value, α is the ratio of the unshielded H₂ dissociation rate to H₂ formation rate, and G is an average H₂ self-shielding factor in dust clouds. The product αG is expressed as:

$$\alpha G = 0.59 I_{UV} \left(\frac{100 \text{ cm}^{-3}}{n_{\text{H}}} \right) \left(\frac{9.9}{1 + 8.9 \tilde{\sigma}_g} \right)^{0.37} \quad (15)$$

where n_{H} is the cloud hydrogen density and I_{UV} is the intensity of radiation field expressed in units of the Draine field.

Thanks to the simultaneous detection of very strong H I H₂ along the quasar J1513+0352 together with dust reddening measurement and the detection of star-formation at small impact parameter (only ~ 1.4 kpc), we could use these expression for the first time at high redshift. The observed H I column density if found to be consistent with the theoretical prediction for a ratio of UV radiation intensity to hydrogen density, $\alpha G \simeq 2$. This constraint complements that from the excitation of C I the UV intensity-hydrogen density plane and implies that the UV field is several times that of our Galaxy. This is consistent with the average UV field estimated from the observed star formation rate and qualitatively consistent with the high excitation of H₂ in the cloud. In other words, we showed that the observed $N(\text{H I})$ together with a measurement of dust reddening or depletion measurement is able to provide important constraints on the physical conditions in a molecular cloud by comparing with the H I-H₂ transition theory. We note in passing the abundances of atomic, molecular gas and dust were derived independently from each other (not using empirical total gas to dust relations). This work is presented in a paper by my current PhD student (Ranjan et al. 2018).

3.7.5 Cloud length-scale

Knowing both the column density and the volumic density of the gas, it becomes easy to derive the typical sizes of the molecular clouds. These are found to have pc-scales, i.e., much smaller than the kpc scales estimated for the warm atomic gas. Such scales are also consistent with the dimensions of molecular clouds seen in our Galaxy. This leads to an interesting observational feature: the size of the H₂ absorbing clouds means that they are actually comparable with the expected size of the emission line regions of the background quasar. In other words, the light beam from the quasar may not be considered at a true point source (i.e. infinitely small) anymore since the absorbing cloud may cover only partly the background source. This means that quasar light may leak around the absorber, resulting in a shift of the apparent zero-flux level. We have shown a clear evidence for this effect for the $z_{\text{abs}} = 2.34$ system towards Q1232+082 ($z_{\text{qso}} = 2.57$), but found evidence for partial coverage in a few other systems as well (Albornoz Vásquez et al. 2014; Klimenko et al. 2015; Krogager et al. 2016b).

3.7.6 Cloud structure

While the excitation of H_2 and C I provide very useful diagnostics of the average physical conditions, numerical modelling of the cloud, reproducing self-consistently the abundance and excitation of atomic and molecular species can provide more in depth study of these conditions and, in particular, investigate how these may vary inside the cloud. Two main numerical codes are available for these studies, the Meudon PDR code (Le Petit et al. 2006) and Cloudy (last described in Ferland et al. 2013).

For example, we used numerical modelling to interpret the the fine-structure levels of neutral oxygen (detected for the first in an intervening DLA) and those of singly ionised silicon (2^{nd} detection) in an extremely strong DLA (Noterdaeme et al. 2015b). These species have different sensitivities to the excitation processes and help to further understand the physical conditions. We found that single-phase photo-ionisation models fail to reproduce simultaneously all observed quantities and suggest that the cloud has a stratified structure, with H_2 found not only in a cold and dense phase ($T \sim 80$ K, $n \sim 300 - 1000 \text{ cm}^{-3}$), but also in a slightly warmer phase (~ 250 K). These contain a fraction of the total H I column density, and the cloud is likely embedded in an even warmer phase (>1000 K), contributing strongly to the excitation of O I . This suggests again that the cloud could be located in a high-pressure region. It could actually correspond to a photo-dissociation region close to a bright star where temperatures are raised in the illuminated side.

In the case of the high-metallicity system towards J0000+0048, already presented above, our set of observational constraints is very extended, with both high-resolution observations (at $R \sim 70\,000$) and multi-wavelength medium resolution observations. These allowed us to measure accurate column densities of various ionised, neutral and molecular species (including H_2 , HD and CO) but also measure the amount of dust and its associated extinction law thanks to the UV to IR coverage. Through numerical modelling with Cloudy, we could reproduce the wide set of observables fairly well with a single cloud illuminated by UV photons. The internal structure of the cloud, as derived from our model is illustrated on Fig.22. We found that the line of sight probes regions deeper than the H I-H_2 transitions, and that CO is found in the cloud interior. Our model also required small dust grains and a high ionisation rate by cosmic rays, which maintain the molecular fraction less than unity, even after the H I-to-H_2 transition. We also found that small dust grains are needed to explain the molecular abundances. The presence of small grains is also in agreement with the observed steep extinction curve that also features a 2175 \AA bump. All this suggests that a simple metallicity-scaling may not be appropriate to understand the conversion from H I to H_2 , in particular at high redshift, and that the actual properties of dust grains must be taken into account.

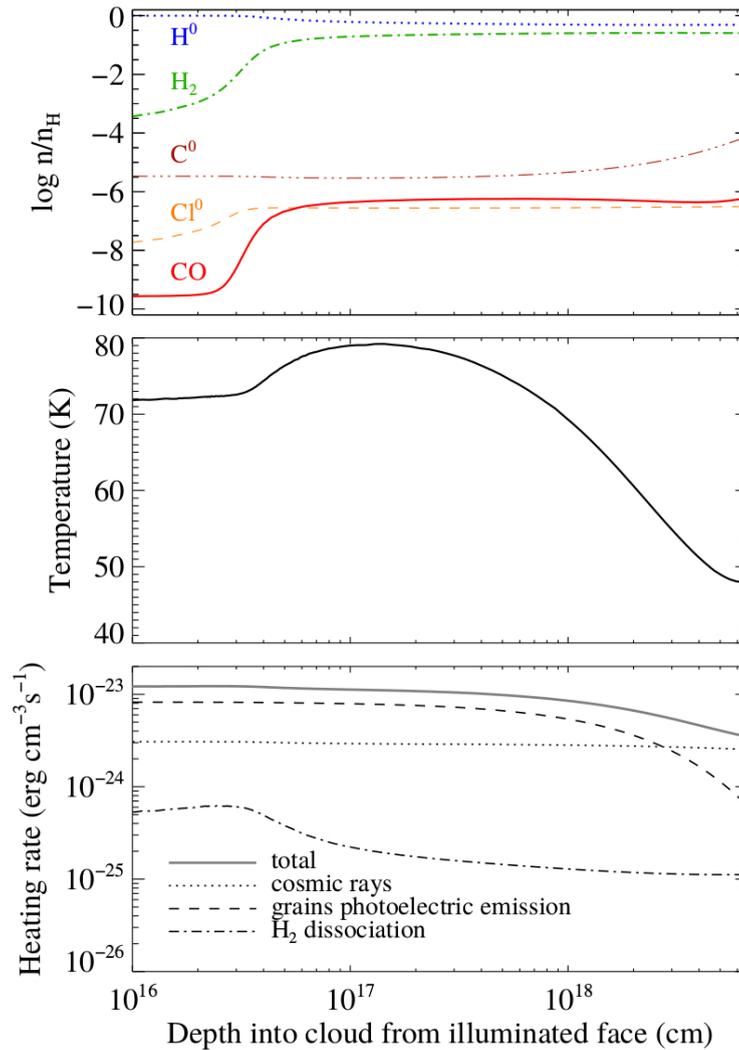


Figure 22: Internal structure of the molecular cloud detected at $z = 2.53$ towards the quasar J0000+0048. The top panel represents the abundances of various species as a function of the depth into the cloud. The middle panel shows the variation of the kinetic temperature and the bottom panel the main heating and cooling rates. Cosmic ray heating becomes dominant in the very interior of the clouds and maintains the molecular fraction less than one even deeper than the transition front (corresponding to H_2 self-shielding from UV photons). [From [Noterdaeme et al. 2017](#)]

3.8 Use of molecular lines for fundamental physics and cosmology

Physics and cosmology as we know today have been remarkably successful in reproducing most of the available observations with only a small number of parameters. However, it also requires that 96% of the mass-energy content of the Universe is in mysterious forms (dark energy and dark matter) that has never been seen in the laboratory. This shows that our canonical theories of gravitation and particle physics may be incomplete, if not incorrect. Improving the sensitivity of current observational constraints is therefore of utmost importance, irrespective of whether it is consistent with the current standard physics –in which case it will reject other scenarios– or whether it will instead favour new physics. The detection of molecular species at high-redshift can provide original and sensitive constraints on these topics.

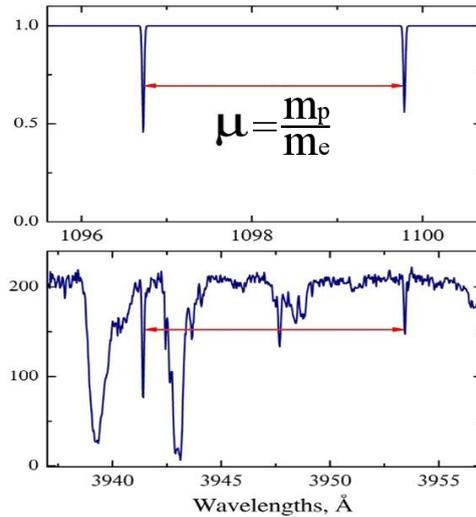


Figure 23: Illustration of quasar absorption lines to constrain the variation of fundamental constants with cosmic time. The relative wavelengths of lines depend on the fundamental constants so that a change over cosmological time-scales results in a difference between the relative wavelength in the laboratory (top) and those measured at high redshift (bottom).

3.8.1 Variation of fundamental constants

The rest-frame wavelengths of molecular lines depend on the proton-to-electron mass ratio (μ) with a sensitivity K_i that is calculated for each line. The observed wavelength of a given H_2 line i is written as

$$\lambda_i^{\text{obs}} = \lambda_i^{\text{lab}} (1 + z_{\text{abs}}) \left(1 + K_i \frac{\Delta\mu}{\mu} \right), \quad (16)$$

where $\Delta\mu/\mu \equiv (\mu_z - \mu_{\text{lab}})/\mu_{\text{lab}}$ is the relative difference between the proton-to-electron mass ratio measured in the absorbing system, μ_z , and in the laboratory, μ_{lab} . For a single line, the observed λ_i^{obs} obviously does not allow to obtain z_{abs} and $\Delta\mu$ simultaneously. However, the *relative* wavelengths of several lines arising from the same gas does not depend on z_{abs} anymore but only on $\Delta\mu$ since their sensitivities K_i are all different. The idea, illustrated in Fig. 23, is thus very simple. In principle, it is sufficient to compare the relative wavelengths of two or more H_2 lines observed at high-redshift with those obtained in the laboratory to get a constraint on $\Delta\mu$ over a timescale that corresponds to the redshift of the absorber, i.e. several Gyrs.

In practice, the wavelength measurements must be very precise and accurate and several measurements (possibly at different redshifts) are needed to constrain the results. Systems with a large number of observed H_2 lines, simple velocity profile and non-saturated lines are preferred. This is the case for example for the systems at $z_{\text{abs}} = 2.402$ and 2.65 towards respectively HE 0027–1836 (Noterdaeme et al. 2007a) and J0643–5041 (Noterdaeme et al. 2008a), for which we obtained stringent constraints (Rahmani et al. 2013; Albornoz Vázquez et al. 2014), but other systems were also analysed by different groups. We found that wavelength distortions can mimic a change in μ and must be taken into account in the error budget. The quest for constraining fundamental constants is therefore also very useful to understand the different systematics that can affect the accuracy of wavelength measurements. Consequently, this is also useful to make progress in the instrumental designs and to understand the probed astrophysical sites. It is also important to explore different ways to constrain μ by, for example, using different species to uncover possible unknown systematics. We tentatively applied the technique for the first time to HD (Noterdaeme et al. 2008a) but the data quality was not high enough for the constraint to be competitive with those obtained from H_2 . Another possibility is to compare the redshift of absorption lines in the UV with the 21-cm absorption. Such studies provides actually

a constraint on $x \equiv g_p \alpha^2 / \mu$, where g_p is the proton gyromagnetic factor and α is the fine structure constant. We have discussed the possible differences in the optical and radio lines of sight and used two systems with compact radio emission (unresolved at milli-arcsecond scales) to obtain $\Delta x/x = (0.2 \pm 1.6) \times 10^{-6}$. This translates to $\Delta\mu/\mu = (0.0 \pm 1.5) \times 10^{-6}$ if we assume α and g_p have not varied (Rahmani et al. 2012).

The electronic transition lines of carbon monoxide can also be used to constrain μ . The first such study was performed by Daprà et al. (2016) using the molecular absorber at $z_{\text{abs}} = 2.69$ toward SDSS J1237+0647 (Noterdaeme et al. 2010a). Instead of fitting each individual line independently, the method considers that, within a CO band, the line strengths and widths are tied by physics. The column densities are assumed to follow a single excitation temperature, significantly reducing the number of free parameters, and the Doppler parameter is ascribed to the turbulent motions in the absorber, rather than representing a kinetic temperature. I then collaborated with this team to apply the same technique to the system towards J0000+0048. The constraint we obtained from this study, $\Delta\mu/\mu \approx (1.8 \pm 2.2_{\text{stat}} \pm 0.4_{\text{syst}}) \times 10^{-5}$ (Daprà et al. 2017) is however not as stringent as those already obtained from H₂ in other systems, but demonstrates the feasibility of the technique.

Currently, the different studies point toward a relative change of μ that is consistent with zero, down to a few parts per million over Gyrs time-scales (Ubachs et al. 2016, and references therein). Progress in the future are bound to the spectrograph capabilities and the availability of molecular absorbers. Bright quasars are naturally preferred to achieve high S/N ratios, but most quasars are actually faint. This is even accentuated for molecular absorbers that are frequently associated to dust, hence extinguishing further the quasar light. New stable, high-resolution instruments with excellent wavelength calibration on telescopes with large collecting area are therefore needed to make further progress. As remarked by Petitjean et al. (2009), in order to prepare for such observations, it remains fundamental to identify more molecular absorption systems. I have presented efficient ways to do so in Sect. 3.1, Sect. 3.3, Sect. 3.2. A summary of currently known and published high- z molecular absorption systems with follow-up observations is shown in Table. 2. We finally note that we have now identified a large number of strong H₂ systems directly in the SDSS-IV, but which are not yet followed-up at higher spectral resolution.

3.8.2 Evolution of the cosmic microwave background temperature

The prediction and discovery of the cosmic microwave background (CMB) radiation is a milestone of modern cosmology. The standard hot Big-Bang theory predicts that, if gravitation is described by general relativity and electromagnetism by Maxwell theory, then photons propagate along null geodesics and the CMB black-body temperature must follow the relation $T_{\text{CMB}}(z) = T_{\text{CMB}}(z = 0) \times (1 + z)$. Any departure from this relation would have strong implications (Uzan et al. 2004) either indicating a violation of the hypothesis of local position invariance (and thus of the equivalence principle) or that the number of photons is not conserved –with the constraint that the energy injection does not induce spectral distortion of the CMB. In the first case, this should be associated with a variation of the fundamental constants. In the second case, it can be associated with decaying dark-energy (e.g. Lima et al. 2000; Jetzer et al. 2011) or with axion-photon-like coupling process (Jaeckel & Ringwald 2010). Direct measurements of the CMB temperature-redshift relation therefore provides an important test of fundamental physics and cosmology.

The CMB temperature at $z = 0$ is known with high accuracy from direct observations of its black-body spectrum ($T_{\text{CMB}}(0) = 2.72548 \pm 0.00057$ K, Fixsen 2009). Two methods have been identified so far to measure T_{CMB} at $z > 0$. The first one relies on measuring a weak spectral distortion of the CMB due to the Sunyaev-Zel'dovich (S-Z) effect in the hot intra-cluster gas, that depends on T_{CMB} at the cluster's redshift. *Planck* data has recently allowed precise measurements over $0 < z < 1$, but with an accuracy degrading with

increasing redshift (from $\Delta T \sim 0.05$ K at $z \sim 0.2$ to $\Delta T \sim 0.6 - 1.0$ K at $z \sim 0.9$). At higher redshift, it is possible to derive T_{CMB} by measuring the excitation of interstellar atomic or molecular species that have transition energies in the sub-millimetre range and can be excited by CMB photons. When the relative population of the different energy levels are in radiative equilibrium with the CMB radiation, the excitation temperature of the species equals that of the black-body radiation at that redshift. Therefore, the detection of these species in diffuse gas, where collisional excitation is negligible, provides one of the best thermometers for determining the T_{CMB} in the distant Universe (Bahcall & Wolf 1968).

Carbon monoxide is an interesting species to perform such measurements. Unlike H_2 , CO has a permanent dipolar momentum allowing direct radiative excitation. In addition, the energy differences between the rotational levels of carbon monoxide (CO) are close to kT_{CMB} at high redshift, making this molecule a CMB-thermometer of choice. From the first detection of CO, we constrained the CMB temperature to be $T_{\text{CMB}}(z = 2.42) = 9.15 \pm 0.7$ and showed that collisional excitation of the molecule should be marginal (Srianand et al. 2008b). I expanded the sample with new CO detections, and, parametrising the departure from the linear relation as $T_{\text{CMB}}(z) = T_{\text{CMB}}(0)(1+z)^{1-\beta}$ (Lima et al. 2000, for models with decaying dark energy), we derived $\beta = -0.007 \pm 0.027$ (Noterdaeme et al. 2011). The accuracy on β has been improved since then by a factor of about two, by combining our CO-based measurements with ATCA sub-mm observations of interstellar species in the peculiar lensing galaxy at $z = 0.89$ towards PKS 1830-211 (Muller et al. 2013) and new S-Z data at $z < 1$ ($\beta = 0.006 \pm 0.013$ by Hurier et al. (2014) and $\beta = 0.005 \pm 0.012$ by Saro et al. (2014)). The different observational constraints are shown on Fig. 24, along with several theoretical relations.

More recently, we observed a CO excitation temperature of $9.9^{+0.7}_{-0.6}$ K in the very molecular rich system towards SDSS J0000+0048 (Noterdaeme et al. 2017). Thanks to our detailed modelling of the physical conditions (see Sect. 3.7), we were able to correct for a small contribution of collisional excitation, that increases the observed excitation temperature by 0.3 K compared to the CMB alone. This system highlights the fact that, to make further progress in getting accurate $T_{\text{CMB}}(z)$ measurements from interstellar species, collisional excitation needs to be taken into account (see also Sobolev et al. 2015). Maeder (2017) even proposed that the temperature excess due to collisional excitation is actually large so that high- z measurements based on interstellar species could actually be consistent with a model in scale-invariant cosmology. We note however that the corresponding $T_{\text{CMB}}(z)$ (dashed purple line in Fig. 24) would then be inconsistent with the SZ measurements as well. In addition, we are currently reinvestigating the excitation of CO with the help of numerical modelling (Klimenko et al, in prep) and do not find large corrections as proposed by Maeder, in particular when restricting to the first few rotational levels, as we actually did (see e.g. Noterdaeme et al. 2010a). Still, it is true that our ability to correctly estimating the contribution of non-CMB excitation processes will be the main limiting factor in the future, when new generation of instruments should yield very precise CO column density measurements in different rotational levels (red dot on Fig. 24). It is therefore important to pursue our investigations of the physical conditions in the gas. Conversely, given the success of standard cosmology in explaining a wealth of observations, it would actually be very surprising that the CMB temperature does not follow the expected relation. The observed excitation of CO rotational levels could then provide a way to constrain the local physical conditions in the gas.

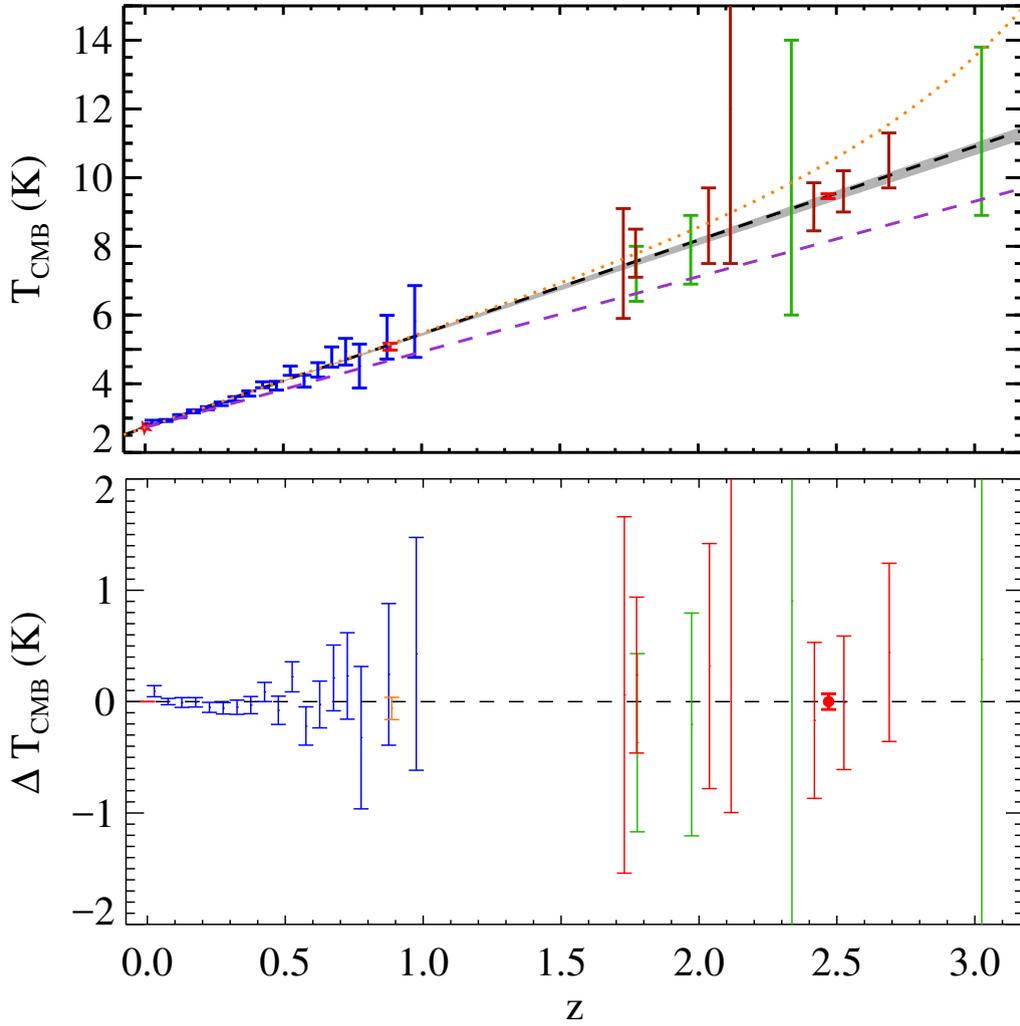


Figure 24: Black-body temperature of the cosmic microwave background radiation as a function of redshift (top) with the difference with respect to the standard adiabatic $T(z) = (1+z)T(0)$ relation (black dashed line) shown in the bottom panel. The red star represents the measurement at $z = 0$ (Fixsen 2009). Blue points correspond to Planck measurements based on the S-Z effect (Hurier et al. 2014). The orange error bar corresponds to the updated measurement at $z = 0.89$ towards PKS1830–211 from the excitation of molecules observed in the sub-mm (Muller et al. 2013). Measurements at higher redshift come from the excitation of atomic carbon (green, Ge et al. 1997; Cui et al. 2005; Srianand et al. 2000; Molaro et al. 2002) and carbon monoxide (brown, from Noterdaeme et al. (2011), with two new points from Ma et al. (2015, at $z \approx 2.12$, with large error bars) and Noterdaeme et al. (2017, at $z \approx 2.53$, after correcting for collisional excitation). The grey shaded area correspond to the combined constraints on the convenient (but not unique) $T_{\text{CMB}}(z) = (1+z)^\beta$ parametrization (Lima et al. 2000). To highlight the need of high- z measurements, we show that using more general parametrizations (e.g. Bassett & Kunz 2004), it is possible to find models that deviate strongly at high- z (orange line) while being consistent with the measurements at $z < 1$. The model with scale-invariant cosmology by Maeder (2017) is shown as dashed purple line. The red point with small error bar at $z \sim 2.5$ is an example of simulated measurement that should be achievable with very high resolution spectrograph on forthcoming extremely large telescopes.

Table 2: Published molecular absorbers at $z > 1.5$ with medium/high resolution follow-up observations¹

Quasar	z_{abs}	$\log N(\text{H I})$	[M/H]	$\log N(\text{H}_2)$	$\log N(\text{HD})$	$\log N(\text{CO})$	References ²
J0000+0048	2.525	$20.80^{+0.10}_{-0.10}$	$+0.46^{+0.45}_{-0.45}$	$20.44^{+0.03}_{-0.03}$	$16.64^{+0.31}_{-0.21}$	$14.95^{+0.05}_{-0.05}$	Not17
Q0013-0029	1.973	$20.83^{+0.05}_{-0.05}$	$-0.59^{+0.05}_{-0.05}$	$18.86^{+1.14}_{-1.14}$	Ge97,Pet02
B0027-1836	2.402	$21.75^{+0.10}_{-0.10}$	$-1.63^{+0.10}_{-0.10}$	$17.30^{+0.07}_{-0.07}$	Not07b,Rah14
J0203+1134	3.387	$21.26^{+0.07}_{-0.08}$	$-1.25^{+0.10}_{-0.10}$	$15.60^{+0.77}_{-0.77}$	Sri12
B0347-3819	3.025	$20.73^{+0.05}_{-0.05}$	$-1.17^{+0.07}_{-0.07}$	$14.53^{+0.06}_{-0.06}$	Lev02,Led03
B0405-4418	2.595	$21.05^{+0.10}_{-0.10}$	$-1.12^{+0.10}_{-0.10}$	$18.14^{+0.07}_{-0.10}$	Led03
B0528-2505	2.811	$21.35^{+0.07}_{-0.07}$	$-0.91^{+0.07}_{-0.07}$	$18.22^{+0.11}_{-0.12}$	$13.22^{+0.02}_{-0.02}$...	Lev85,Kli15
Q0551-3638	1.962	$20.70^{+0.08}_{-0.08}$	$-0.35^{+0.08}_{-0.08}$	$17.42^{+0.45}_{-0.73}$	Led02
J0643-5041	2.659	$20.95^{+0.08}_{-0.08}$	$-0.91^{+0.09}_{-0.09}$	$18.54^{+0.01}_{-0.01}$	$13.65^{+0.00}_{-0.01}$...	Not08,Alb14
J0812+3208	2.626	$21.35^{+0.10}_{-0.10}$	$-0.81^{+0.10}_{-0.10}$	$19.93^{+0.05}_{-0.05}$	$15.70^{+0.07}_{-0.07}$	< 12.81	Jor09,Bal10
J0816+1446	3.287	$22.00^{+0.10}_{-0.10}$	$-1.10^{+0.10}_{-0.10}$	$18.66^{+0.20}_{-0.16}$	Gui12
J0843+0221	2.787	$21.82^{+0.11}_{-0.11}$	$-1.52^{+0.08}_{-0.10}$	$21.21^{+0.02}_{-0.02}$	$17.35^{+0.15}_{-0.34}$	< 13.40	Bal15
J0857+1855	1.730	$19.70^{+0.20}_{-0.20}$	$-0.19^{+0.21}_{-0.21}$	$13.54^{+0.05}_{-0.05}$	Not11
J0917+0154	2.107	$21.00^{+0.07}_{-0.07}$	$-0.18^{+0.12}_{-0.12}$	$20.11^{+0.06}_{-0.06}$...	< 14.07	Not18
J0918+1636	2.580	$20.96^{+0.05}_{-0.05}$	$-0.12^{+0.05}_{-0.05}$	$17.60^{+1.45}_{-1.45}$	Fyn11
J1047+2057	1.775	$20.58^{+0.05}_{-0.05}$	$-0.18^{+0.12}_{-0.12}$	$14.40^{+0.07}_{-0.07}$	Not11
J1117+1437	2.001	$19.80^{+0.10}_{-0.10}$	$+0.35^{+0.14}_{-0.14}$	~ 18	...	< 13.13	Not18
Q1232+0815	2.338	$20.90^{+0.08}_{-0.08}$	$-1.35^{+0.12}_{-0.12}$	$19.57^{+0.10}_{-0.10}$	$15.52^{+0.17}_{-0.17}$	$12.55^{+0.00}_{-0.00}$	Ge01,Iva10,Bal11
J1237+0647	2.690	$20.00^{+0.15}_{-0.15}$	$+0.34^{+0.12}_{-0.12}$	$19.21^{+0.13}_{-0.12}$	$14.48^{+0.05}_{-0.05}$	$14.17^{+0.09}_{-0.09}$	Not10b
J1311+2225	3.092	$20.75^{+0.10}_{-0.10}$	< -0.24	$19.69^{+0.01}_{-0.01}$...	< 13.43	Not18
B1331+0170	1.777	$21.17^{+0.07}_{-0.07}$	$-1.22^{+0.04}_{-0.04}$	$19.71^{+0.07}_{-0.07}$	$15.03^{+0.13}_{-0.11}$...	Cui05,Bal10,Car11
J1337+3152	3.174	$21.36^{+0.10}_{-0.10}$	$-1.45^{+0.22}_{-0.22}$	$14.09^{+0.03}_{-0.03}$	Sri10
J1439+1118	2.418	$20.10^{+0.10}_{-0.08}$	$+0.16^{+0.11}_{-0.11}$	$19.38^{+0.10}_{-0.10}$	$14.87^{+0.03}_{-0.03}$	$13.89^{+0.02}_{-0.02}$	Sri08a,Not08b
J1443+2724	4.224	$20.95^{+0.10}_{-0.10}$	$-0.63^{+0.10}_{-0.10}$	$18.29^{+0.07}_{-0.07}$	Led06
B1444+0126	2.087	$20.25^{+0.07}_{-0.07}$	$-0.80^{+0.09}_{-0.09}$	$18.16^{+0.14}_{-0.12}$	Led03
J1456+1609	3.350	$21.70^{+0.10}_{-0.10}$	$-1.32^{+0.11}_{-0.11}$	$17.10^{+0.09}_{-0.09}$	Not15
J1513+0352	2.464	$21.83^{+0.02}_{-0.02}$	$-0.84^{+0.23}_{-0.23}$	$21.31^{+0.02}_{-0.02}$	~ 16.40	< 13.60	Ran18
J1604+2203	1.641	$14.59^{+0.11}_{-0.11}$	Not09a
J1646+2329	1.998	$19.75^{+0.10}_{-0.10}$	$+0.13^{+0.18}_{-0.18}$	$18.02^{+0.11}_{-0.11}$...	< 13.40	Not18
J1705+3543	2.038	$20.62^{+0.12}_{-0.12}$	$+0.07^{+0.14}_{-0.14}$	$14.14^{+0.03}_{-0.03}$	Not11
J2100-0641	3.091	$21.05^{+0.15}_{-0.15}$	$-0.73^{+0.15}_{-0.15}$	$18.76^{+0.03}_{-0.03}$	$13.83^{+0.06}_{-0.06}$...	Bal15,Iva15,Jor10
J2123-0050	2.059	$19.18^{+0.15}_{-0.15}$	$+0.12^{+0.15}_{-0.15}$	$17.94^{+0.01}_{-0.01}$	$13.84^{+0.20}_{-0.20}$	< 13.07	Mal10,Tum09,Mil10
J2140-0321	2.340	$22.40^{+0.10}_{-0.10}$	$-1.05^{+0.13}_{-0.13}$	$20.13^{+0.07}_{-0.07}$...	< 13.73	Not15
J2225+0527	2.131	$20.69^{+0.05}_{-0.05}$	$-0.09^{+0.05}_{-0.05}$	$19.40^{+0.10}_{-0.10}$	Kro16b
J2257-1001	1.836	$20.40^{+0.07}_{-0.07}$	$+0.03^{+0.08}_{-0.08}$	$19.50^{+0.10}_{-0.10}$...	< 13.09	Not18
B2318-1107	1.989	$20.68^{+0.05}_{-0.05}$	$-0.85^{+0.06}_{-0.06}$	$15.49^{+0.03}_{-0.03}$	Not07b
J2331-0908	2.143	$21.15^{+0.15}_{-0.15}$	$-0.54^{+0.15}_{-0.15}$	$20.57^{+0.05}_{-0.05}$...	$13.65^{+0.03}_{-0.03}$	Not18
J2336-1058	1.829	$20.33^{+0.02}_{-0.02}$	$-0.22^{+0.10}_{-0.10}$	$19.00^{+0.12}_{-0.12}$...	< 12.93	Not18
J2340-0053	2.055	$20.35^{+0.15}_{-0.15}$	$-0.74^{+0.16}_{-0.16}$	$18.20^{+0.20}_{-0.20}$	Jor10
B2343+125	2.431	$20.40^{+0.07}_{-0.07}$	$-0.87^{+0.10}_{-0.10}$	$13.69^{+0.09}_{-0.09}$	Pet06,Not07b
B2348-0108	2.426	$20.50^{+0.10}_{-0.10}$	$-0.62^{+0.10}_{-0.10}$	$18.52^{+0.29}_{-0.49}$	Pet06,Not07a

¹ We have now identified a large number of additional strong H₂ systems in the SDSS but which are not yet observed at higher spectral resolution. ² References are abbreviated as three letters of first author+year: Albornoz Vázquez et al. (2014); Balashev et al. (2010, 2011, 2015); Carswell et al. (2011); Cui et al. (2005); Fynbo et al. (2011); Ge et al. (1997, 2001); Guimaraes et al. (2012); Ivanchik et al. (2010); Jorgenson et al. (2009, 2010); Klimenko et al. (2015); Krogager et al. (2016b); Ledoux et al. (2002, 2003, 2006b); Levshakov & Varshalovich (1985); Levshakov et al. (2002); Malec et al. (2010); Milutinovic et al. (2010); Noterdaeme et al. (2007a,b, 2008a,b, 2009a, 2010a, 2011, 2015b, 2017, 2018); Petitjean et al. (2002, 2006); Rahmati & Schaye (2014); Ranjan et al. (2018); Srianand et al. (2008b, 2010, 2012); Tumlinson et al. (2010).

4 Dust

Most of the photo energy released by star-formation is absorbed by dust grains present in the interstellar medium and re-radiated in the infrared. In addition, the presence of dust strongly influences the physical state of the gas through photo-electric heating, UV shielding, and formation of molecules on the surface of grains. Therefore the knowledge of the amount, physical properties and composition of dust at high redshift is of great importance to understand galaxy formation and evolution. While the study of the dust is almost inextricable from that of atomic and molecular gas and has therefore already been mentioned previous sections, I intent here to give here a light on the discussion from the perspective of dust. In addition, I discuss the effects of dust on the overall census of quasar absorbers.

4.1 Depletion

As already mentioned in Sect. 2.3.2, several metal species tend to lock into dust grains. This means that the observed column density (i.e. in the gas phase) of a given species is related to the true total column density as $N(X)_{\text{obs}} = N(X)_{\text{int}} - N(X)_{\text{dust}}$, where $N(X)_{\text{dust}}$ is the column density removed from the gas phase. The simplest approach to estimate the abundance of dust is then to compare the observed abundance of a refractory and a volatile element, assuming an intrinsic abundance ratio, generally taken to be Solar.

There are however two issues with this method. First, volatile species can themselves present some degree of depletion so that their observed abundance doesn't correspond to the intrinsic value. For example, while sulphur is generally considered undepleted and frequently used as reference zero-point, it has recently been found to be depleted by a factor of about two in the very nearby ISM towards α Leo (Gry & Jenkins 2017). Second, the intrinsic abundance ratio may not be Solar, in particular in galaxies at high redshift since it depends on the actual chemical enrichment history. Indeed stars of different masses produce elements with different yields and over different time-scales. These two issues can however be alleviated by studying statistically the relative abundances of various species over a wide range of values. For example, Wolfe et al. (2005) argued that the flattening of the observed silicon-to-iron ratio at low [Si/H] values indicates a nucleosynthetic ratio $[\text{Si}/\text{Fe}] \approx 0.3$ and that the increase of [Si/Fe] at high metallicity results from increasing depletion of iron. More recently, De Cia et al. (2013) studied altogether the observed abundance ratio of several species as a function of [Zn/Fe] and derived the corresponding depletion sequences. They found that [Cr/Zn] decreases as a function of [Zn/Fe], indicating more and more depletion of chromium when that of iron also increases. Since the slopes of the relation between two abundances ratios are different, this allows one to remove the need of a perfectly undepleted species and derive the true metallicity from the observed abundance of at least two species. For example, the depletion of zinc is actually not zero and increases with metallicity, but this increase is slower than that of iron. This means that a correction depending on [Zn/Fe] can be applied to the observed [Zn/H] to obtain the true intrinsic metallicity, and, at the same time, the abundance of dust. These authors further showed that nucleosynthesis is not a major effect on the observed abundance ratios.

From our survey for H_2 in DLAs, we remarked that the probability to detect H_2 in DLAs increases with the overall DLA metallicity (Petitjean et al. 2006). This suggested already that dust plays an important role, if metallicity correlates with dust content (Ledoux et al. 2003), but it may also be related to the galaxy mass, if the relation between velocity-width and metallicity is indeed due to an underlying mass-metallicity relation (Ledoux et al. 2006a). From our VLT/UVES survey, we observed a neat increase in the detection rate of H_2 with the depletion factor (Noterdaeme et al. 2008b). Since the detection rate depends on the H_2 column density while the depletion factor is a *local* gas property (not integrated), we further compared $N(\text{H}_2)$ with $N(\text{Fe})_{\text{dust}}$, the column density of iron locked into dust grains (i.e. missing from the gas phase). We found that

all H₂ systems are detected in systems with $N(\text{Fe})_{\text{dust}} > 5 \times 10^{14} \text{ cm}^{-2}$. There are however a few systems above this limit but that don't have H₂ down to low detection limits ($\sim 10^{15} \text{ cm}^{-2}$). This could be due to the line of sight passing through gas with low volume density (or a concatenation of clouds), but with a long total path, so that it ends up building a relatively high total integrated dust column density without ever reaching the local conditions for efficient H₂ formation.

4.2 Reddening of the background quasar light

The presence of dust along the line of sight produces a reddening of the background source's light. Indeed, dust grains absorb and scatter short wavelength photons more than they do with long wavelength photons, resulting in change in the shape of an observed spectrum. How much change is occurring as a function of wavelength is called the extinction law. The exact shape of this law depends on the dust grain size distribution and their chemical composition. Indeed, to a first approximation, individual dust grains absorb photons with wavelengths smaller than their physical size. The fact that absorption occurs more in the UV, less in the optical and very little in the infra-red already tells us that there must be many more small grains than large grains. Some features can also be present, the most remarkable of which being an increased absorption over a range of wavelength centred at 2175 Å (Stecher 1965, called the 2175 Å bump), which is thought to arise from carbonaceous grains and/or a mix of large polycyclic aromatic hydrocarbon (PAH) molecules (Draine 2003). Typically, the dust grain size distribution is parametrised by a truncated single power law with the form $dn/da \propto a^{-q}$, where n is the number of grains with radius in the range $[a, a + da]$. Nozawa & Fukugita (2013) have shown that the range of extinction curves seen in the Milky-Way and Small Magellanic Cloud can be reproduced with a mixture of graphite and silicates with a power law index $q \approx 3.5 \pm 0.2$ with cut-off at small grain radii $a_{\text{min}} = 0.05 \mu\text{m}$ (albeit not well constrained by their data) and at large grain radii, $a_{\text{max}} = 0.2 - 0.3 \mu\text{m}$.

Extinction measurements are typically obtained by comparing a pair of stellar spectra of the same spectral type, one of which being reddened and the other one not (see Fitzpatrick & Massa 2007). This technique has been applied to different regions of the Milky-Way but also of the Magellanic clouds. While there are still large variations from region to region, average extinction curves for the Milky-Way, the LMC (and LMC2 supershell) and SMC (in order of decreasing bump strength) are frequently used as convenient references for comparison (Gordon et al. 2003).

At high redshift, one can use a similar approach. The observed spectrum of a reddened quasar is compared with a composite "unreddened" quasar spectrum. By studying the spectral slopes of a large sample of SDSS quasar spectra with detected Mg II absorbers and comparing with a matching sample (with similar redshift and *i*-band magnitude) without absorber, York et al. (2006) found $\langle E(B - V) \rangle \sim 0.01$ with a SMC-like extinction law. Similarly, several authors have attempted to measure the average dust extinction of DLAs, using different techniques, comparing the photometric colours of quasars with/without DLAs (Vladilo et al. 2008; Fukugita & Ménard 2015), their spectral index (Murphy & Liske 2004; Murphy & Bernet 2016), or comparing composite spectra (Frank & Péroux 2010; Khare et al. 2012). While there is some discrepancy between the results, these can be accommodated within the systematic uncertainties due to the different techniques and sample selections (in terms of e.g. redshift or $N(\text{H I})$ column density). The main point is that all studies indicate very small average reddening, with $E(B - V)$ of the order of a few milli-magnitudes only. On average, the dust content of atomic gas appears therefore to be low in the distant Universe but consistent with a metallicity scaling of the dust-to-gas ratio.

Notwithstanding, there are unambiguous evidences for significant dust reddening associated to cold gas and molecular gas. At low redshift, both diffuse interstellar bands and the 2175 Å bump have been detected in

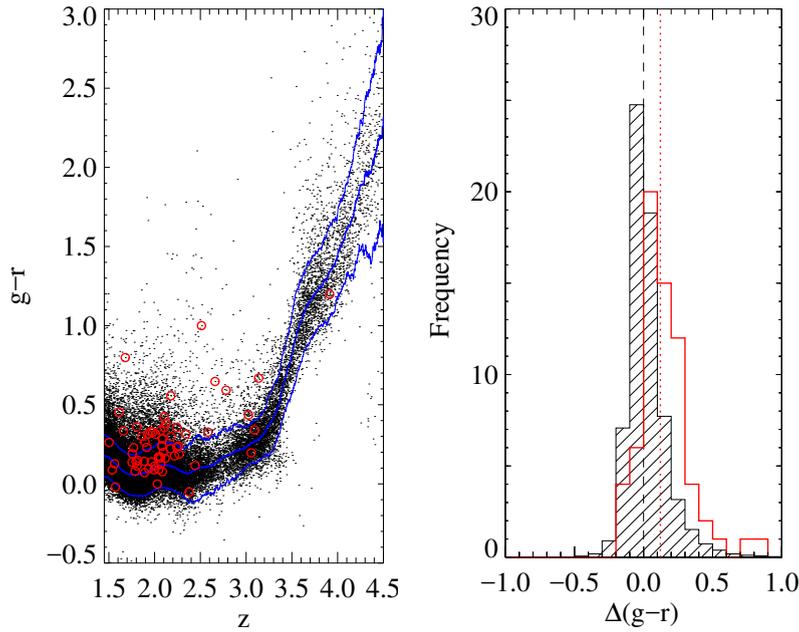


Figure 25: Left: $g - r$ photometric colours of quasars with C I-absorbers in the SDSS-II (red) compared to the overall population of quasars (black points) as a function of redshift. The blue line indicate the average and dispersion of the $g - r$ values. Right: Distribution on the colour excess ($\Delta g - r$), after subtracting this redshift-dependent average value. Clearly, the colour-distribution of quasars with and without C I absorbers are very different, quasars with C I-absorber having on average $g - r$ higher by 0.1 mag. [From Ledoux et al. 2015]

the $z_{\text{abs}} = 0.524$ 21-cm absorption system towards AO 0235+164, with $E(B - V) = 0.23$ (Junkkarinen et al. 2004). In our survey for 21-cm at intermediate redshifts ($1.1 < z < 1.45$) we also identified two systems with strong prominent 2175 Å bump and $E(B - V) \sim 0.3$ (Srianand et al. 2008a). Soon after this discovery, we also found a strong reddening and a prominent 2175 Å bump in a CO-bearing absorber at $z = 1.64$ (Noterdaeme et al. 2009b). From our complete sample of C I systems in the SDSS, we found that the colours of quasars with intervening C I absorbers are significantly redder from the rest of the population (see Fig. 25), with a colour excess of the order of $\Delta g - r \sim 0.1$ mag. The relatively short wavelength coverage of SDSS however leaves some degeneracy between reddening, intrinsic quasar brightness and choice of extinction law. These issues are now solved thanks to the UV-to-NIR wavelength coverage of the Xshooter spectrograph. We have measured reddening of the order of $E(B - V) \sim 0.1$ (Zou et al. 2018), with a range of extinction laws, with and without 2175 Å bump. This work, lead by our PhD student Siwei Zou, also focused on the relation between dust and the Ca II and Na I lines, only made possible at high-redshift thanks to the UV to NIR capabilities of Xshooter. She showed that the the rest-frame lines equivalent widths (for a given dust content) are typically larger at high redshift than at low redshift. This indicates that the widths is likely due to a large kinematic spread, i.e., the high- z systems may be perturbed due to, e.g. strong gas flows or interactions.

Instead of trying to match the observed spectrum with a quasar composite reddened by a given extinction law, it is also possible to derive the extinction law when flux-calibrated data is available over a wide range of wavelengths. This is the case for the $z_{\text{abs}} = 2.5$ molecular absorber towards J0000+0148 (Fig. 26). The derived extinction curve is similar to that seen towards ζPer, with a similar 2175 Å feature but a steeper slope in the UV. This indicates a relative increased amount of small grains. It happened from numerical modelling that extending the typical dust-grain size distribution down to a cut-off smaller than the canonical value not only reproduces well the observed extinction curve, but also explains the observed abundances of neutral and molecular species (Noterdaeme et al. 2017). A key point is that small dust grains give a high total surface for a given total mass abundance of dust (or equivalently amount of metals missing from the gas phase, i.e.

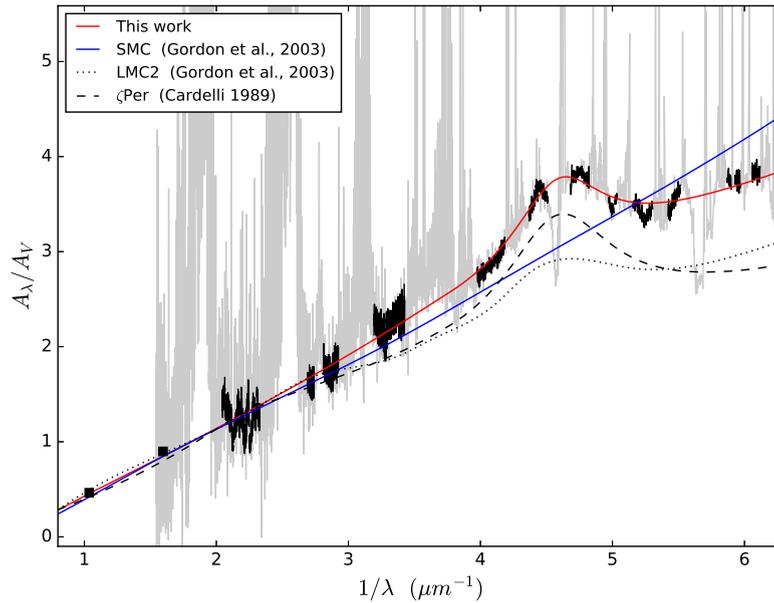


Figure 26: Extinction curve towards the quasar J0000+0048 as a function of inverse wavelength in the DLA rest-frame, normalised to the V -band. The grey line shows the X-shooter spectrum divided by the quasar template of [Selsing et al. \(2016\)](#), with the regions used in the fit (without absorption or emission line) highlighted in black. The two black squares indicate the photometry in the K and $W1$ bands. The red solid line shows our best-fit extinction curve, which present a clear bump at 2175 \AA . SMC (solid blue) and LMC2 (dotted) curves from [Gordon et al. \(2003\)](#) are shown for comparison. The dashed extinction curve corresponds to a MW extinction law with the single parameter, R_V , set to the value towards ζPer ([Cardelli et al. 1989](#)). [From [Noterdaeme et al. \(2017\)](#)]

the depletion factor). This increases the shielding efficiency for a given dust-to-gas mass ratio and provides a larger surface for the formation of molecules. This opens very interesting routes to pursue investigation: We already know that the conversion from atomic to molecular gas depends strongly on the amount of dust, but the grain size distribution of the latter seems to play also an important role.

4.3 Dust bias

Since the presence of dust along the line of sight to a quasar alters the colours of the later, and since quasars are typically selected in the optical for their blue colours, it has soon noted that quasars could be missed in current samples due to dust-biasing ([Pei et al. 1991](#); [Boissé et al. 1998](#)). Of course, there is little doubt that large column densities of dust can redden and extinguish the background quasars so much that they will be missed, but the question is to know to which extent the dust bias will affect the absorbers statistics and hence our view of gas in the Universe.

Since the typical reddening associated to atomic gas traced by DLAs is very small, it has often been considered that the census of the atomic gas is not strongly affected. This could be questioned however since many of the reddening studies were based on DLAs detected towards optically-selected quasars. By construction, such samples therefore did not include a population of dusty DLAs. However, the average observed reddening is almost zero, well below the limit beyond which a DLA would be missed. This means that there cannot be a significant population of moderately dusty DLAs that would otherwise make the average reddening higher. One possibility is that the distribution of dust is bi-modal with a population of very dusty DLAs (being missed) and a population of dust-free DLAs. From studying radio-selected quasars, [Ellison et al. \(2001\)](#) and [Jorgenson et al. \(2006\)](#) concluded that the incidence of DLAs is not significantly affected by dust, albeit larger

samples are need to draw a firm conclusion. On the other hand, we do have evidences for highly reddened systems. This means that, if not bimodal, the dust distribution could simply be peaked at low values but with a long tail extending to large reddening. In that case, the H I census should also not be strongly affected, but that of metals, dust, and molecules could be more biased in magnitude-limited samples.

[Pontzen & Pettini \(2009\)](#) have made a first estimate of the effect of dust-obscuration on absorbers statistics using a Bayesian parameter estimation analysis that includes radio and optical observations of DLAs. Their study then considers the fraction of quasars that is extinguished beyond the magnitude limit of SDSS and found that Ω_{HI} is only affected by a 7% only, while the total cosmological density of metals, Ω_m , could be underestimated by a factor of two. We are currently re-investigating this issue by considering the full quasar selection function of SDSS-II, meaning that the consequence of colour shifts is also taken into account rather than just a limiting magnitude for spectroscopic observations. We also considered updated metallicity and $N(\text{HI})$ distributions. Our preliminary results indicate an effect of the order of 20% for the H I census. This is significant compared to the statistical uncertainties but it has little influence on the discussion of the co-evolution of atomic gas and star-formation (SFR changes with redshift are over a much larger dynamical range). We also expect that the H I census should be less biased in the SDSS-III (BOSS) and SDSS-IV (eBOSS) surveys, since the quasar selection has been improved significantly and the magnitude limits for spectroscopic observations have been relaxed. On the other hand, we confirm the biasing of the metal/dust census, that one should keep in mind when discussing the metallicity evolution as well as the incidence of cold and molecular phases as a whole.

In order to further investigate the issue of dusty biasing, in particular on the detection of cold gas and molecular phases, we have been developing two strategies. The first approach is to perform a tailored search for quasars that are reddened out of the current optical SDSS selection. I have been collaborating with the team of Johan Fynbo to develop a selection based on optical plus infrared photometry (UKIDSS and then WISE). In a series of papers ([Fynbo et al. 2013](#); [Krogager et al. 2015, 2016a](#)), we showed that our method efficiently identifies red quasars. However, it turn out that most of them were actually red because of their intrinsic properties rather than the presence of dust in a foreground absorber. Still, the selection allowed us to identify several absorbers with relatively high column densities of metals and dust (e.g. [Krogager et al. 2016b](#); [Fynbo et al. 2017](#); [Heintz et al. 2018](#)), in a way independent from those described in Sects. 3.1, 3.2 and 3.3. An interesting possibility is to select systems upon the presence of a 2175 Å bump. This allows one to disentangle much better the origin of the dust (i.e. intrinsic to the quasar or intervening), since the 2175 Å bump can be assigned a redshift (which is hard to do for the sole reddening). I have collaborated with the team of Jian Ge to systematically search for such systems using SDSS spectra (e.g. [Ma et al. 2015](#); [Zhang et al. 2015](#)). We could also consider selecting 2175 Å absorbers directly from photometric data, since the shift of the quasar in colour space should be distinguishable from the main quasar and stellar locus. We are currently investigating this possibility as well.

The other approach is to obtain bias-free statistics of cold gas by selecting quasars independently on the optical colours. This is the strategy we have adopted for the MeerKAT Absorption Line Survey (PIs Gupta and Srianand), for which we have been allocated 2000 h of observing time to blindly search for 21-cm absorption at $0 < z < 1.5$ (see [Gupta et al. 2016](#)). For this project, we are selecting compact radio sources on the basis of a single flux limit at 1.4 GHz (>200 mJy). In order to increase the probability that the sources are actually high-redshift quasars, we further considered the infra-red colours from the WISE all-sky survey ([Wright et al. 2010](#)). Spectroscopic observations of the selected candidates, with the South African Large Telescope and the Nordic Optical Telescope, indicate that about 75% are indeed quasars (for NOT-based results, see [Krogager et al. 2018a](#)). The 21-cm observations with MeerKAT have started in 2018 and should provide exciting results in the near future.

5 Star-formation

An important question that I have not discussed until now is how absorption systems seen in quasar spectra actually relate to the population of galaxies seen in emission. On one hand, we have shown that absorption spectroscopy allows to investigate the distribution, chemical and physical properties of the different gas phases, which also gives indirect clues on the present and past star-formation. On the other hand, over the past two decades, astronomers have found efficient observational strategies to detect and study galaxies in emission in the distant Universe (e.g. [Le Fèvre et al. 2015](#)). However, connecting the gas properties to that of the underlying galaxy population remains observationally very challenging, in particular at high redshift. Because absorption systems are selected against background sources, they are selected based on the gas cross-section, regardless of the luminosity of the associated object. Quasar absorbers are therefore expected to sample the luminosity function of galaxies over a wide range of luminosities (e.g. [Cen 2012](#)). In addition, the total cross-section of DLAs is much larger than that of starlight-emitting regions in observed galaxies, meaning that a large fraction of DLAs potentially probe atomic gas in the halo or circum-galactic environments, as suggested by high-resolution simulations of galaxy formation (e.g. [Pontzen et al. 2008](#)). This must be tested by observations. In this section, I will summarize our contribution to recent progress in the field, at low and intermediate redshifts (Sect. 5.1) and at high redshifts (Sect. 5.2).

5.1 Low and intermediate redshifts

5.1.1 Connection between metal absorbers and galaxies

At low and intermediate redshifts ($z < 1$), it is in principle easier to learn about the galaxy population giving rise to absorption systems than it is in the very distant Universe. The connection has first been investigated for Mg II systems (e.g. [Bergeron 1986](#)), since the corresponding doublet ($\lambda\lambda 2796, 2800$) is redshifted in the optical range and easily spotted in quasar spectra. At that time, there was actually still a debate to know whether absorption systems were due to intervening galaxies or were associated to the quasar. The works by [Bergeron \(1986\)](#) and [Bergeron & Boissé \(1991\)](#) demonstrated that galaxies were found at the absorber's redshift confirming the intervening galaxy hypothesis. Galaxies responsible for quasar absorption systems were subsequently found to span a range of galaxy types (e.g. [Steidel et al. 1994](#); [Le Brun et al. 1997](#); [Rao et al. 2003](#); [Chen et al. 2005](#)). and the emerging picture was that the gas traced by Mg II can be spread over large distances (several tens of kilo-parsecs) from the host galaxy. Indeed, since singly-ionised magnesium is expected to arise from both ionised and neutral gas (with a wide range of H I column densities), the Mg II doublet can trace a variety of environments, including galactic winds. Notwithstanding, the success rate of detecting Mg II absorption lines in the spectrum of quasars with known foreground galaxies is much less than unity ([Bechtold & Ellingson 1992](#); [Bowen et al. 1995](#); [Tripp & Bowen 2005](#)), suggesting that the gaseous halos around galaxies is not uniformly populated. [Wild et al. \(2007\)](#) later focused on Ca II absorbers, which, according to their metal and dust properties, are likely to probe neutral gas with H I column densities greater than the nominal limit for DLAs ([Wild et al. 2006](#)). From stacking the SDSS DR3-DR4 quasar spectra shifted at the redshift of the absorber, these authors detected the average nebular line emission falling into the fibres and concluded that the star-formation rate per unit area of DLAs is an order of magnitude below the expectation from the Kennicutt-Schmidt relation. These authors then rejected the possibility that the underlying galaxy population could contribute significantly to the overall SFR density at low redshifts.

Based on a large statistics of Mg II systems in the SDSS-DR7, we showed for the first time a statistical correlation between the equivalent width (W_r) of the Mg II absorption and the strength of the [O II] and [O III]

emission lines found super-imposed to the quasar spectra (i.e. collected within the aperture of the SDSS fibre centred on the quasar, see [Noterdaeme et al. 2010a](#)). The much lower average emission compared to that of individual detections suggests that faint galaxies contribute appreciably to the population of strong Mg II systems at intermediate redshifts. The interpretation later given by [Ménard et al. \(2011\)](#) that the correlation between equivalent width and observed [O II] flux represents a correlation between equivalent width and star-formation may in turn be too simplistic since fibre losses are at play and must be considered ([López & Chen 2012](#)). It therefore appears premature to use Mg II systems directly as a probe of the global evolution of star-formation.

We took the opportunity of different fibre sizes used for SDSS-II (3 arcsec diameter) and SDSS-III (2 arcsec) to investigate further this issue ([Joshi et al. 2017b, 2018](#)). We searched for nebular emission in individual spectra and actually found no strong correlation between $Wr(\text{Mg II})$ and [O II] fibre-luminosity when restricting to narrow redshift ranges, while the observed [O II] luminosity was found to depend strongly on redshift. This is consistent with the expected luminosity evolution of galaxies but we also showed that the finite fibre sizes plays a very crucial role this dependence, since the angular separation between the quasar line of sight and the centroid of the galaxy decreases with increasing redshift.

In order to shed different light on the absorber-galaxy connection, we also used a reversed approach by first searching for nebular [O III] emission lines superimposed to quasar fibre spectra and then characterising the absorption properties ([Noterdaeme et al. 2010a](#)). We uncovered a population of low luminosity [O III] galaxies that would have not been targeted on their own in SDSS (as galaxy target) would a background quasar not being present. Despite the line of sight passing through the star-light emitting regions of the foreground galaxies, the observed dust-extinction remains modest. We further increased the sample by searching for oxygen and/or Balmer lines over a wider range of redshifts ([Straka et al. 2015](#)). From subtracting the quasars from the images, we found that the galaxies span a range of masses ($\log M_*/M_\odot = 7.3 - 11.5$) with properties comparable to field galaxies in the same redshift range. We also observed an anti-correlation between dust extinction and impact parameter. All this suggests that lines of sight passing at small galacto-centric distances may be more subject to dust-biasing and hence under-represented in quasar absorption studies.

5.1.2 Searching for 21-cm absorption in or around known galaxies

We have used 21-cm absorption to investigate the cold gas properties of known intervening galaxies in a series of papers ([Gupta et al. 2013](#); [Srianand et al. 2013](#); [Dutta et al. 2017](#); [Gupta et al. 2018](#)). The idea was to search for 21-cm along quasar lines of sight passing through or close to intervening galaxies seen in emission. Interestingly, we found a variety of situations, with some examples given below:

We reported the 21-cm detection in two quasar-galaxy pairs at $z \sim 0.3$ with, in both cases, estimated column densities of H I larger than a few times 10^{20} cm^{-2} . In the first pair, the galaxy is an early-type galaxy at impact parameter of about 14 kpc, which could mean that a reservoir of cold H I gas with low-level of star-formation activity can be found in the outskirts of luminous red galaxies. In the second pair, the galaxy is found from its emission lines on top of the quasar and produces a strong reddening of the background quasar. Despite of this, we detected only a weak 21-cm absorption. This could mean that a large fraction of the gas is warm or that the covering factor of radio source by the foreground gas is small. Another possibility is that the H I column density is much less than suggested by the extinction, which would imply that the dust-to-gas ratio in this system is much higher than seen locally.

In another case, the background quasar is piercing through the spiral disc at an impact parameter of about 4 kpc from the centroid ([Srianand et al. 2013](#)). We detected a strong 21-cm absorption and diffuse interstellar bands (DIBs), as well as a strong reddening of the background source. Using VLBA spectroscopy, we demonstrated that the cold gas is structured at the parsec scale.

Finally, by studying the 21-cm statistics quasars with known foreground galaxies with a range of impact parameters, we found that the covering factor of 21-cm decreases from $C_{21} \approx 0.24$ at $\rho < 15$ kpc to $C_{21} \approx 0.06$ at $\rho \sim 15 - 35$ kpc and depends on the azimuthal angle between galaxy axis and location of the background quasar (Dutta et al. 2017). The higher 21-cm detection rate in $z < 1$ DLAs compared to lines of sight selected by the presence of a galaxy at small impact parameter indicates that the H I gas has a small covering factor and a patchy distribution around low redshift galaxies.

5.2 High redshifts

Characterising the absorber-galaxy connection is more difficult at high redshift and attempts to do so have often been frustrated by the difficulty of detecting galaxies at the low end of the luminosity function, and/or close to the bright quasar. I here summarise current efforts and progress made thanks the combined power of statistics, new powerful instruments and optimised observational strategies.

5.2.1 Statistical detection of Ly- α emission

The wide core of the damped Lyman- α system removes the quasar flux over several angstroms, leaving free access to the detection of Ly- α emission from the intervening galaxy down to a null impact parameter. By coadding quasar fibre spectra at the DLA rest-frame, it is then possible to constrain the average Ly- α emission that enters the fibre. Using 341 DLAs from the SDSS-II catalogue (Noterdaeme et al. 2009a), with $\log N(\text{H I}) > 20.6$ (so that the wings of the DLA do not contaminate the zero-flux core at the SDSS resolution), we constrained the average Ly- α luminosity emitted within 1.5 arcsec (or 12 kpc) to be less than 2×10^{41} erg s $^{-1}$ (Rahmani et al. 2010). The resulted sensitivity was already better than most surveys for Ly- α emitters at that time and comparable with the very deep long slit spectroscopic observations by Rauch et al. (2008). We concluded that DLAs arise from the low luminosity end of the population of Ly- α emitting galaxies, in which case they contribute little to the overall SFR density. Alternatively, it is possible that the associated galaxies are statistically at impact parameters larger than the SDSS fibre radius (12 kpc). This second possibility is substantiated by the detection of O VI absorption in the composite spectrum, which could trace highly ionised gaseous winds.

We then extended this work to a larger sample drawn from the SDSS-III DLA catalogue (Noterdaeme et al. 2012b) and detected Ly- α emission with a characteristic double-peak profile (Joshi et al. 2017a). We showed that the emission cannot be explained by Ly- α fluorescence resulting from ionisation equilibrium with the UV meta-galactic background alone and instead requires in-situ star-formation. We found that the strength of the emission depends on the equivalent width (W_r) of metals. Since W_r depends both on the column density and the velocity extend, the higher Ly- α luminosity associated with systems with higher W_r could indicate enhanced star-formation in high-metallicity systems and/or an easier escape of Ly- α enabled by outflowing gas.

We also statistically investigated systems with very high H I column densities (Noterdaeme et al. 2014). These are expected to be connected both spatially and physically to star-forming regions in galaxies if a Schmidt-Kennicutt law applies to quasar absorbers, as proposed by Chelouche & Bowen (2010). If the column density is high enough, then we can expect most galaxies to fall within the projected radius of the SDSS fibre removing the degeneracy between impact parameter and star-formation rate (also in Sect. 5.1). Simulations (e.g. Pontzen et al. 2008; Rahmati et al. 2013) indicate that at $\log N(\text{H I}) \sim 22$, the impact parameters reduce to a few kpc with a small dispersion. By co-adding the spectra of about a hundred extremely strong DLAs, we indeed detected a significant Ly- α emission. We found that both the average emission and its distribution

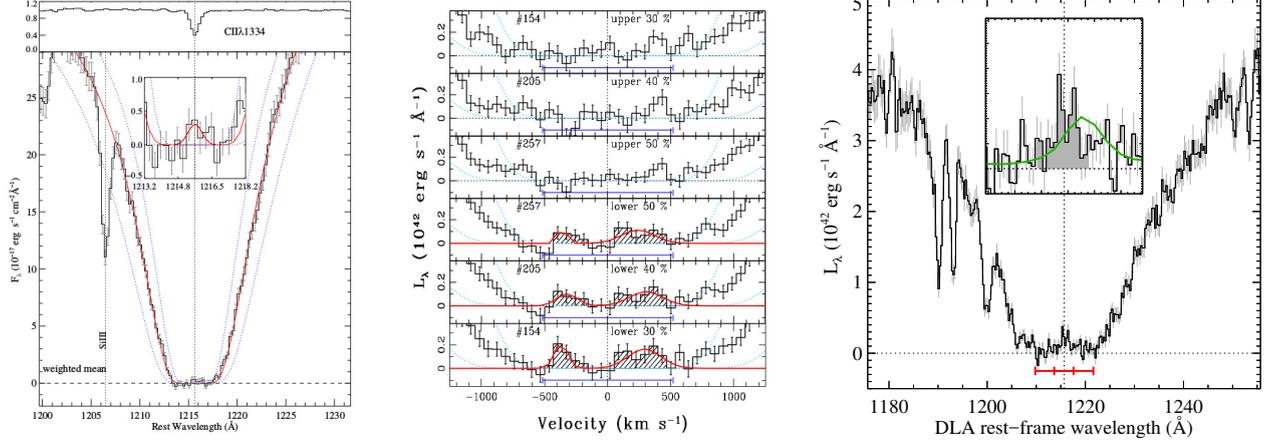


Figure 27: Statistical detection of Ly- α emission in the core of DLAs. The different panels correspond to stacked DLA spectra from different works. Left: Our early proof of concept using SDSS-II and $\log N(\text{H I}) > 20.6$ (Rahmani et al. 2010). Centre: Updated stacking using SDSS-III/BOSS DLA spectra with $\log N(\text{H I}) > 21$ and $W_r(\text{Si II } \lambda 1526) > 0.4 \text{ \AA}$ (Joshi et al. 2017a) in sub-samples of decreasing dust reddening from top to bottom. Right: A stronger Ly- α emission is seen in extremely strong DLAs with $\log N(\text{H I}) > 21.7$ (Noterdaeme et al. 2014). The green curve in the inset zoom shows the scaled average profile for Lyman-break galaxies from Shapley (2011).

corresponds to that expected for the population of Ly- α emitting galaxies (LAEs) with $L(\text{Ly-}\alpha) > 10^{41} \text{ erg s}^{-1}$. In addition, the properties of the Ly- α emission in terms of apparent velocity extent and dependence on dust are also similar to this population of LAEs. By comparing the incidence of extreme DLAs with the luminosity distribution function of LAEs (Cassata et al. 2011), we derived the mean physical projected extent of the high column density gas. Indeed, integrating the high $N(\text{H I})$ -end of the H I-distribution function, we obtain the incidence of ESDLA per unit absorption distance (dN_{ESDLA}/dX) which we can express in units of co-moving distance as

$$\frac{dN_{\text{ESDLA}}}{dl_c} = \frac{dN_{\text{ESDLA}}}{d\chi} \frac{H_0(1+z)^2}{c} \approx 2 \times 10^{-6} \text{ Mpc}^{-1}. \quad (17)$$

On the other hand, the comoving incidence of LAEs that give rise to ESDLAs is

$$dN_{\text{LAE}}(L > L_{\text{min}})/dl_c = \sigma_{\text{gas}}(1+z)^2 d\Omega \int_{L_{\text{min}}}^{\infty} \Phi(L) dL, \quad (18)$$

where $\sigma_{\text{gas}} = \pi r_{\text{gas}}^2$ and r_{gas} is the mean physical projected extent of the gas with $N(\text{H I}) \geq 5 \times 10^{21} \text{ cm}^{-2}$ and $\Phi(L)$ is the Ly- α luminosity function. By equating Eq. 17 and Eq. 18 at $\langle z \rangle = 2.5$ and using our derived value $L_{\text{min}} = 10^{41} \text{ erg s}^{-1}$, we obtained $r_{\text{gas}} \approx 2.5 \text{ kpc}$. In other words, the expected number density of LAEs within an impact parameter of 2.5 kpc from the quasar lines of sight accounts for the incidence of ESDLAs. This confirms our hypothesis and shows that fibre losses are negligible. Finally, we measured an average star-formation rate that is consistent with the expectation from the Schmidt-Kennicutt law if the typical galaxy radius is 2.5 kpc. Our results therefore support a close association between the population of emission selected LAEs and strong DLA host galaxies.

5.2.2 Individual direct detection

While statistical detection of galaxy light allows to draw broad conclusions about the connection between atomic gas and star-formation, detecting galaxies responsible for individual systems is required to complete the picture and uncover the relations between gas surface density, chemical enrichment (metallicity, dust and molecular content), and the star-formation rate and impact parameter of the associated galaxy. A variety of techniques have been used through the years to attempt overcoming this very challenging task.

Broad-band imaging is in principle best suited to collect light from faint galaxies. However, this technique also collects the photons from the bright background quasar so that it remains hard to disprove the presence of a faint galaxy at a small impact parameter, even using space-based observations and subtracting the quasar PSF. For example, [Warren et al. \(2001\)](#) reported 41 candidate DLA galaxy counterparts around 18 quasars using HST/NICMOS imaging. Subsequent spectroscopic observations of three candidates by [Møller et al. \(2004\)](#) resulted in the detection of Ly- α emission in the DLA core (hence on top of the quasar) and not associated to any of the reported candidates. In order to be sensitive to the galaxy light even at the smallest impact parameters, an alternative, originally proposed by [O’Meara et al. \(2006\)](#) is the so-called double-absorber technique: The galaxy responsible for an absorption system is searched in broad-band imaging in a wavelength range where the flux of the quasar is completely removed by a Lyman-limit system at higher redshift. We tried this technique for two such configurations taking the opportunity of new blue-sensitive CCD on the VLT/FORS1 imager at that time, but this also resulted in non-detections ([Christensen et al. 2009](#)). [Fumagalli et al. \(2015\)](#) also used this technique on a larger sample and find a very low *in-situ* star-formation rate (i.e. right at the location of the absorber).

Narrow-band imaging allows one to search for galaxies in a selected narrow-redshift range but is limited by the available filters and manufacturing a customised filter for a given redshift is costly. By selecting DLAs with redshift corresponding to an existing filter on Subaru, [Kashikawa et al. \(2014\)](#) detected an extended Ly- α emission associated to a sub-DLA system at $z_{\text{abs}} = 3.115$, which they confirmed through follow-up spectroscopy. Unfortunately, little information about the properties of the absorbing gas is available for this system.

Deep (multi-wavelength) spectroscopy makes easier the detection and identification of weak emission lines from the galaxy even on top of a relative bright continuum from the quasar. It also provides directly the redshift of the galaxy, so that it is easier to reject any coincident interloping galaxy. A drawback is the small sky coverage around the background quasar. A strategy we have adopted is then to perform observations with several slit position angles on the sky. Not only this increases the sky coverage around the quasar, but it becomes also possible to measure the impact parameter by triangulation. The arrival of the Xshooter spectrograph at the VLT was a fantastic opportunity to apply this technique: this instrument has a high sensitivity and medium spectral resolution, well suited to observe and resolve galaxy emission lines, and a large wavelength coverage (from UV to NIR) making it possible to simultaneously search for Ly- α oxygen and Balmer emission lines at high redshift. A blind search of galaxies associated to the overall population of DLAs would have been quite inefficient given the expected range in impact parameters and star-formation rates (e.g. [Cen 2012](#)). Instead, we developed two strategies to respectively target (a) galaxies with relatively high star-formation rates and (b) galaxies at small impact parameters:

(a) *Metal-rich systems*: The first approach, coordinated by Johan Fynbo, was to search for galaxy counterparts of high-metallicity DLAs. The idea is that these systems could be associated to more massive (and more luminous galaxies) if the velocity-width-metallicity correlation observed in high- z DLAs by [Ledoux et al. \(2006b\)](#) is the signature of a mass-metallicity relation. We note that the interpretation of the correlation between kinematics and metallicity is debated. Large velocity widths could also be due to winds driven

by star-formation (in possibly lower-mass galaxies), but the neat result is the same: an expected correlation between star-formation and strength of metal absorption lines. Our observing campaign with X-shooter was quite successful and allowed us to detect several DLA galaxies published in a series of papers (Fynbo et al. 2011, 2013; Krogager et al. 2012, 2013). In a recent work (Krogager et al. 2017), we summarised the properties of the total sample and discussed the findings using a simple model that links the luminosity and impact parameter of the galaxy to the metallicity measured in absorption. We showed that the observations do not necessarily follow the model prediction in terms of star-formation and impact parameter for a given metallicity, but that the expected and observed distributions match very well. The object-to-object variations are actually the result of a significant scatter in the underlying relations. Our model also reproduces at the same time the detection rates and results of other surveys based on different techniques: spectral stacking (Sect. 5.2.1), direct imaging (Warren et al. 2001), imaging below the Lyman-limit (Fumagalli et al. 2015) and integral field spectroscopy (Péroux et al. 2012). The model allows to understand better the observations that can be counter-intuitive. Indeed, for a given galaxy, one expects a metallicity gradient, decreasing with galacto-centric distance. However, since metallicity also correlates with mass and mass with galaxy extent, the tendency is actually that the mean impact parameter increases with the DLA metallicity⁸. A further improvement to this model would naturally to include the $N(\text{HI})$ -dependence of the gas cross-section when the model currently only considers metallicity. Similarly, the linkage between gas column density and impact parameter should also be a function of the galaxy luminosity.

(b) *Extreme $N(\text{HI})$ systems*: The second approach, led by myself, is to consider systems at the high end of the column density distribution. Our hypothesis is that the impact parameters in this regime are typically small, of the order of a few kpc only or less when the column density is of the order of 10^{22} cm^{-2} , as indicated by our stacking results (Noterdaeme et al. 2014, Sect. 5.2.1). This is also seen in 21-cm emission maps of galaxies in the local Universe (Zwaan et al. 2005) and in simulations for the high- z Universe (e.g. Pontzen et al. 2008; Rahmati & Schaye 2014). In addition, the skewed $N(\text{HI})$ distribution of DLAs associated to GRB afterglows compared to that of intervening DLAs (e.g. Fynbo et al. 2009) is also consistent with our picture since GRB are linked to the death of massive stars and should statistically probe more central (gas-rich) regions of galaxies than random (quasar) lines of sight crossing neutral gas.

Our first candidate, observed both with the MagE spectrograph on the Magellan telescope and with X-shooter indeed resulted in the detection of the DLA galaxy exactly on top of the background quasar, with an impact parameter between the emission centroid and the line of sight less than 1 kpc (Noterdaeme et al. 2012c). We detected Lyman- α , oxygen and Balmer emission lines. The Ly- α line is double-peaked and spatially resolved with "blue" and "red" photons originating from slightly different locations. The absorption and emission data is best interpreted as the observations of a young, gas-rich, compact starburst galaxy, from which material is expelled through collimated winds powered by a vigorous star formation activity. We substantiated this picture by modelling the radiative transfer of Ly- α photons in the galactic counterpart.

We have now assembled a sample of extremely strong DLAs, all observed with Xshooter, which serves two main science objectives: the study of the microphysics of the HI -to- H_2 transition at the small scales (see Sect. 3.2) and the global relationship between the properties of the galaxy properties and that of the gas *inside* the galaxies. This sample makes the basis for the thesis of my current PhD student (Adarsh Ranjan). From this sample, we already simultaneously detected coincident atomic gas, molecular gas and star-formation at the same place (impact parameter $\rho \sim 1 - 2 \text{ kpc}$ Ranjan et al. 2018). For the first time, we could apply the microphysics-based HI - H_2 transition theories using constraints from the local abundances and excitation of atomic and molecular species together with those obtained on the star-formation activity from the detection of the galaxy in emission. At least two new detections of associated galaxy emission are found among the first 10

⁸The gas absorption cross-section, and hence the probability P_i to be intercepted by a random line of sight, increases as $P_i \propto \rho^2$, where ρ is the impact parameter to the galaxy centroid.

objects analysed, both of which with impact parameter less than a couple of kpc. Since our observations are also sensitive to larger impact parameters (~ 8 kpc perpendicular to the slit, and >30 kpc parallel to the slit), the non-detection in the remaining systems suggests that the corresponding galaxies simply remain below the detection limit.

Integral field spectroscopy is a promising technique since it presents both the advantages of continuous and wide sky-coverage and simultaneous spectroscopy, allowing an extended search for emission lines around the quasar (Péroux et al. 2011; Bouché et al. 2012; Wang et al. 2015). Yet, the wavelength range covered is generally large enough to study the gas properties and cover multiple emission lines so that additional observations are required (when X-shooter does all at once). In addition, the detection rate of blindly selected samples remains small. Progress is however expected in the near future with state-of-the-art instruments like MUSE on the VLT, in particular with the narrow field mode that has just been made available to the community.

5.2.3 Summarising the DLA-galaxy detections: linking impact parameter to gas column density and metallicity

We summarise the DLA-galaxy detections obtained so far at high redshift in Table 3 and compare the observed impact parameter with those expected from simulations (Rahmati & Schaye 2014) as a function of H I column density (see Fig. 28). Overall, the decreasing impact parameter with increasing $N(\text{H I})$ is clearly seen both in the observational data and in the simulations. Similarly, the dispersion also decreases with increasing $N(\text{H I})$. However, the observed data points populate mostly the upper ranges of the impact parameters predicted by simulations, even considering the large expected dispersion (the shaded area represents only the 68%). For example, all the observed impact parameters for systems with $N(\text{H I}) < 5 \times 10^{21} \text{ cm}^{-2}$ are located above the median expected value. One possibility is that the simulations predict too small H I extent around high- z galaxies. However, before asking simulations to make substantial progress, one should understand the selection function of the observed data points and the associated biases. First of all, the DLA-galaxy detections come from heterogeneous searches performed with different techniques, but also different selections. Indeed, only our study of extremely strong DLAs is based on the H I column only, and indeed matches very well the simulations. Most of the remaining systems were actually selected upon their high metallicity. These are likely associated to more massive galaxies with larger gaseous extent, and hence statistically large impact parameters. It is well possible that gas from a faint galaxy in a group is responsible for the absorption along the quasar line of sight, but it will be the brightest galaxy in the group (i.e. the tip of the iceberg) that will be most easily detected, and hence at a larger distance.

The two systems with with largest impact parameters illustrate the need to understand well the selection function: Both DLAs have fairly high metallicities, meaning that the quasar may probe gas associated to the far outskirts of a bright galaxy. The first case is actually a sub-DLA in which we detected H₂, CO and HD molecules. Rudie et al. (2017) argued that the line of sight probes molecular gas that has been carried out by a galaxy outflow. This interpretation would be in agreement with the large velocity width of the metal absorption lines. In the second case, the galaxy has been found from CO(3-2) emission with ALMA, at an impact parameter of 28 kpc (Neeleman et al. 2018). However, it is hard to disprove that these systems are actually associated to fainter galaxies located at smaller impact parameter, but below the detection limit. A nice example of such bias is illustrated by ALMA observations of Q0918+1636 by Fynbo et al. (2018). In this case, the previously identified high-metallicity DLA galaxy counterpart at $\rho \sim 16$ kpc remains below ALMA's detection limit. CO emission is however detected from a massive galaxy at sensibly the same redshift but at a much larger impact parameter of 117 kpc. This galaxy would probably have been considered as associated to the quasar absorber if the fainter galaxy hadn't been identified previously. This means that direct association

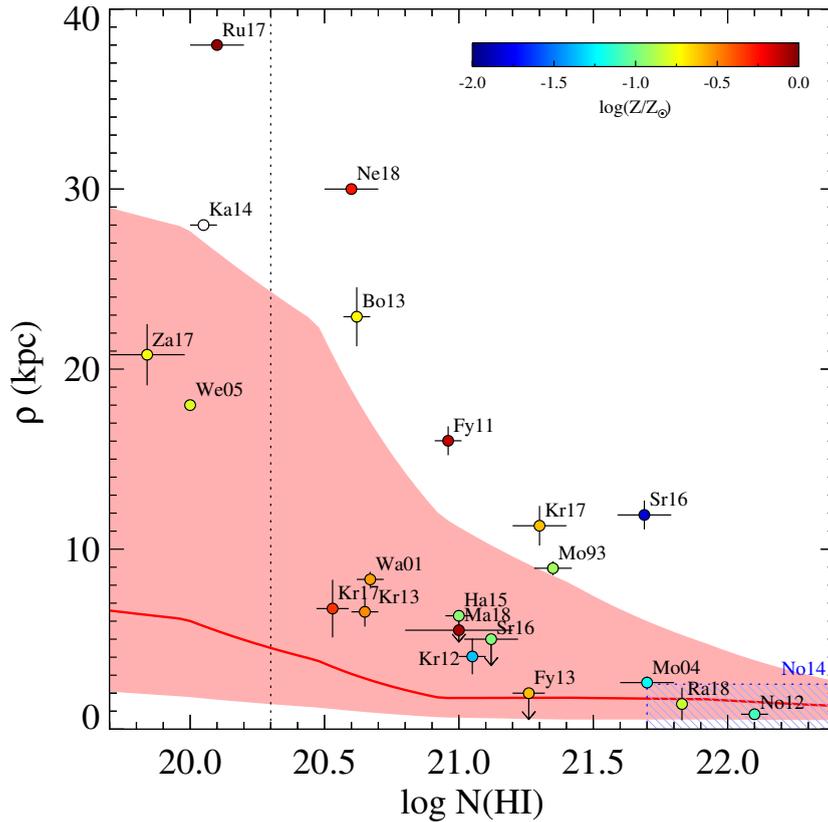


Figure 28: Impact parameter between the centroid of high-redshift galaxies and the line of sight as a function of the atomic hydrogen column density measured along this line of sight. The red line represent the median value in the simulations by [Rahmati & Schaye \(2014\)](#) with the red shaded region encompassing 68% of this impact parameter distribution. Coloured circles represent individual detection of absorbing galaxies, with the colour representing the metallicity measured in the absorbing gas (one system has no metallicity measurement and is filled white). The label indicated next to each point correspond to the reference as indicated in Table 3. The blue hashed region represent the average constraint from DLA stacking ([Noterdaeme et al. 2014](#)). We are currently exploring more in depth this region, where the properties of the gas is probed at very small galacto-centric radius (i.e. a truly interstellar medium, as opposed to circumgalactic medium) with several new detections.

between gas and galaxies should be considered with care. Deeper ALMA observations of the same systems are therefore highly desirable. To be fair, the comparison between simulations and observations should actually be done at a given star-formation detection limit, and taking into account both the gas column density and the metallicity, the latter not being considered in simulations like that of [Rahmati & Schaye \(2014\)](#).

Conversely, it would be very instructive to characterise the absorption properties along quasar lines of sight piercing through or near emission-selected galaxies with known properties, like we already did at low redshift. It is clear that there is still a long way to go to associate absorbing gas and galaxy over a range of properties and as a function of the line of sight location with respect to the galaxies. Different approaches on the relation between gas (including molecular gas) and star-formation will allow us to shed light on a different parts of the overall picture and progress in the field will be made by considering they as complementarity and not substitutive one of the other.

Table 3: Emission counterparts to neutral gas absorption systems at high redshift

Quasar	z_{DLA}	$\log N(\text{H I})$	[M/H]	ρ (kpc)	Ref.
Q1313+1441	1.79	21.30 ± 0.10	-0.60 ± 0.10	11.30 ± 1.10	Krogager et al. (2017, Kr17)
Q2206-19	1.92	20.67 ± 0.05	-0.54 ± 0.05	8.32 ± 0.42	Warren et al. (2001, Wa01)
PKS0458-02	2.04	21.70 ± 0.10	-1.22 ± 0.10	2.59 ± 0.33	Møller et al. (2004, Mo04)
Q1135-0010	2.21	22.10 ± 0.05	-1.10 ± 0.08	0.83 ± 0.08	Noterdaeme et al. (2012c, No12)
Q2059-0528	2.21	21.00 ± 0.05	-0.96 ± 0.06	< 6.3	Hartoog et al. (2015, Ha15)
Q0338-0005	2.22	21.05 ± 0.05	-1.36 ± 0.07	4.04 ± 0.99	Krogager et al. (2012, Kr12)
Q2243-60	2.33	20.62 ± 0.05	-0.72 ± 0.05	22.91 ± 1.64	Bouché et al. (2013, Bo13)
Q2222-0946	2.35	20.65 ± 0.05	-0.50 ± 0.03	6.52 ± 0.82	Krogager et al. (2013, Kr13)
Q0918+1636	2.41	21.26 ± 0.06	-0.60 ± 0.20	< 2.0	Fynbo et al. (2013, Fy13)
Q2348-011	2.42	20.53 ± 0.06	-0.33 ± 0.08	6.70 ± 1.6	Krogager et al. (2017, Kr17)
Q0918+1636	2.58	20.96 ± 0.05	-0.12 ± 0.05	16.02 ± 0.80	Fynbo et al. (2011, Fy11)
PKS0528-250	2.81	21.35 ± 0.07	-0.91 ± 0.07	8.94 ± 0.39	Møller & Warren (1993, Mo93)
J2358+0149	2.98	21.69 ± 0.10	-1.83 ± 0.18	11.9 ± 0.80	Srianand et al. (2016, Sr16)
J2358+0149	3.25	21.12 ± 0.10	-0.97 ± 0.13	< 5.0	Srianand et al. (2016, Sr16)
J1513+0352	2.46	21.83 ± 0.01	-0.84 ± 0.23	1.4 ± 0.90	Ranjan et al. (2018, Ra18)
J1211+0833	2.12	21.00 ± 0.20	-0.07 ± 0.21	5.5	Ma et al. (2018, Ma18)
B1228-113	2.19	20.60 ± 0.10	-0.27 ± 0.10	30	Neeleman et al. (2018, Ne18)
2dF2239-2949	1.83	19.84 ± 0.14	> -0.75	20.8 ± 1.70	Zafar et al. (2017, Za17)
Q1439+1117	2.42	20.10 ± 0.10	-0.03 ± 0.12	38	Rudie et al. (2017, Ru17)
J0310+0055	3.12	20.05 ± 0.05	-	28	Kashikawa et al. (2014, Ka14)
J2236+1325	3.15	20.00	-0.80 ± 0.24	18	Weatherley et al. (2005, We05)

6 Concluding remarks and perspectives

During many years, absorption studies of the neutral gas in the distant Universe were mostly confined to discussions about the number statistics of DLAs, abundance measurements (including still on-going discussions about depletion and nucleosynthesis, e.g. [Vladilo et al. 2018](#)) and semi-quantitative analysis of the gas kinematics. In addition to reaching precision measurements, the presence and properties of molecular gas, dust and associated star-formation has now been added to this discussion.

Large quasar spectroscopic surveys, the availability of (blue-) sensitive instruments on large telescopes and the development of original tools to analyse and interpret both large databases and follow-up data have all played fundamental roles to this end. Indeed, the global census and evolution of absorbers can be studied with excellent statistical power thanks to the many-fold increase in the number of known quasars (mostly from the Sloan Digital Sky Survey). In addition, the large number of surveyed lines of sight has permitted to uncover absorbers arising from cold and molecular gas phases and investigate these with large telescopes. These phases not only act as a sensitive probe of the physical processes at play in galaxies, but also drive the onset of star-formation and hence participate importantly in setting the way the galaxies evolve. Having said this, the lack of overlap between absorption and emission studies constituted for long an important obstacle to link gas and star-formation in galaxies. Thankfully, our improved selection strategies combined with new powerful instruments have lead to substantial progress in uncovering the links in recent years.

Overall, absorption studies now provide us with a window onto the complex, multiphase interstellar medium at high-redshift. Indeed, while the quasar absorbers number statistics is dominated by warm diffuse phases (due to their large cross-section selection), the gas in high redshift galaxies intrinsically presents the same degree of complexity as what is observed in our own Galaxy. We showed that gas with specific chemical or physical properties at high- z can be efficiently targeted using optimised selections. For example, we showed that high (\sim Solar) metallicity molecular-rich systems are efficiently traced by neutral carbon independently of their H I properties, or that systems at the high end of the H I column-density distribution probe galaxies at small impact parameter, where the gas pressure is high. These may have high molecular content despite typically low metallicities and hence represent CO-dark gas that will remain out of reach through CO emission. In addition, we are now able to obtain quantitative measurements of the local chemical composition and physical conditions of the gas, to investigate the chemistry at play between atomic, ionic and molecular species and study the structure of the gas at very small scales.

The micro-physic processes that drive the local gas properties are also extensively studied theoretically and used to explain both observations of the nearby ISM as well as that of gas at high redshift. These processes are seemingly scaled to the chemical enrichment, at least at the first order. In addition, the gas distribution and dynamics is different at high redshift, where gas flows are important and galaxies not yet well-formed. This means that one should refrain from blindly using local relations between e.g. molecular gas and star-formation, between dust-extinction and gas column, or between CO emission and molecular mass to interpret what is seen at high redshift. Quasar absorption studies are very helpful in probing the relations between the ISM constituents in the sense that the main ingredients, such as the column densities of H I, H₂ and CO are measured directly –not through proxies– and the same is true for the dust extinction and the metallicity –that does not require calibrator. A deep understanding the gas properties and its transformational processes based on first principles rather than phenomenological approaches can also have deep implications on the understanding of the evolution of galaxies, well beyond the already fundamental small-scale processes it represents in the interstellar medium. Just to give an example, most hydrodynamical simulations of galaxy formation define a critical gas density, typically $n_H \sim 10 \text{ cm}^{-3}$, above which star formation is triggered. However, simulations generally do not consider the H₂ phase at all. In other words, the simulations shortcut

the H I-H₂ transition, and even when considered, it is generally done in post-processing. This is certainly very useful to see how H₂ can be distributed (i.e. H₂ as a tracer), but not how it will influence the star-formation (i.e. H₂ as a driver). Other parameters are then fine-tuned to alter the star-formation efficiency and match the observations. This includes feedback processes such as AGNs or supernova winds. Therefore, the magnitude of such feedback processes may not be well estimated if the star-formation law depends also on the local chemical and physical conditions. Of course, this is a challenging task in particular since the scales over which the cold and molecular gas is found is much below the typical resolution of these simulations.

Tremendous progress on our understanding of the molecular content of distant galaxies is being made from CO emission studies, in particular with ALMA (see review by [Combes 2018](#)), and similar progress concerning the atomic gas through 21-cm emission will be possible with SKA. Still, quasar absorption lines can in principle also bring useful information about the *global* relationships between the overall gas and star-formation properties as well, although in a statistical manner and not on individual objects. In order to do so, systematic analyses of large samples must be performed, combining different approaches that target a given gas or galaxy property. At the same time, we should make progress in understanding better the corresponding selection functions. For example, we are only recently starting to quantify the effect of dust biasing on the different populations of absorbers and not only on the cosmological census of diffuse atomic gas.

Finally, I would like to remind that quasars also offer unique probes of fundamental physics and cosmology, including the Big-Bang Nucleosynthesis, the variation of fundamental constants and the evolution of the temperature of the cosmic microwave background radiation. These actually constitute important science cases for several instruments on forthcoming ground and space telescopes (e.g. HIRES on the ELT ([Zerbi et al. 2014](#)), Pollux on LUVOIR ([Bouret et al. 2018](#))).

Going beyond longitudinal information on intervening gas

The observational, modelling and theoretical tools, together with our experience in understanding of the gas properties at high redshift can also be very useful to investigate the immediate environment of the background source itself. At the same time, we can move beyond the inherently one-dimensional nature of absorption lines spectroscopy, using multiple sightlines, extended background sources or also considering their different emission components (which are rich in the case of quasars: accretion disc, broad and narrow emission line regions, host galaxy, scattered Ly- α emission). This is briefly discussed below.

Transverse information

While quasar absorption lines provide information about the gas along the line of sight, it is possible to obtain transverse information by observing two (or more) background sources with close separation on the sky (e.g. [Ellison et al. 2007](#); [Finley et al. 2014](#)). This is particularly useful since it provides direct information on the sizes of the absorbers, which can then be compared to the scales derived from the ratio of the column to volumic densities measured along the line of sight. Naturally, the separation between the lines of sight sets the scale over which a given structure can be probed. Quasars with separations of the order of a few arcmin can be used to study the transverse flux correlation of the Ly- α forest and obtain constraints on the structure of the IGM (e.g. [Coppolani et al. 2006](#)). Smaller separations, of the order of a few arcseconds or less, are useful to study the distribution of the atomic gas around galaxies. For example, [Ellison et al. \(2007\)](#) observed several coincident DLA absorption along two quasar lines of sight with projected separation of about 14 arcsec. Because of the still large physical distance between the sightlines at the absorption redshifts (~ 110 kpc), the authors interpret the absorptions as the signature of gas in groups of two or more galaxies rather than arising from a single large galaxy.

Transverse information can also be obtained from a single background source, when its extent on the sky becomes comparable to that of the foreground absorber (or at least the length-scales over which a given gas phase is found). For example, bright galaxies can in principle be used as background sources. [Cooke & O’Meara \(2015\)](#) have observed an almost complete ($\sim 90\%$) absorption of a damped Lyman- α with $\log N(\text{H I}) \sim 21$ in front of a galaxy and argue that the atomic gas must be continuously spread over at least several kpc^2 , but possibly up to 100kpc^2 . However, the observed "mean" absorption depends actually on the source light distribution (which is not expected to be uniform) and the gas distribution in the foreground absorbing screen, which we showed to be patchy and hence with spatial variations of column densities. The stellar emission of galaxies hosting quasars can also provide transverse information in addition that towards their central AGN. We indeed observed statistically some residual flux in the bottom of intervening damped Ly- α systems, which likely corresponds to leaking far-UV emission from the quasar host ([Cai et al. 2014](#)). This means that the intervening neutral gas is not continuously extended over the whole starlight emitting region of the background quasar host. Galaxies will likely become more frequently used as background sources in the near future. Naturally, obtaining absorption spectra for different regions of the background galaxies will provide more information than the average absorption, but will require sensitive IFUs on future extremely large telescopes. Interestingly, [Lopez et al. \(2018\)](#) have already made original use of the MUSE instrument on the VLT by targeting a bright, giant gravitational arc. This allowed them to map the spatial and kinematic distribution of Mg II absorption of an intervening galaxy at $z \sim 1$, which they observed to be clumpy with an absorption strength decreasing with increasing distance from the galaxy.

For the cold and molecular phases, assuming a typical density $n \sim 100 \text{cm}^{-3}$ and a column density $N(\text{H I}) \sim 10^{20} \text{cm}^{-2}$, one can see that the typical length scale should be parsec to sub-parsec. There is no way this can be resolved at high-redshift. However, transverse constraints can still be obtained from the incomplete absorption of the background source, at specific wavelengths where the size of the source is known or can be constrained independently. We have shown that, in several cases, intervening H_2 clouds do not completely cover the background broad emission line region (e.g. [Balashev et al. 2011](#); [Albornoz Vásquez et al. 2014](#); [Krogager et al. 2016b](#)). This sets the absorber’s extent to be of parsec-scales, in agreement with our indirect longitudinal estimates. Conversely, knowing the properties of the intervening gas allows one to put constraints on the extent of the background source at different wavelengths, hence-by providing important constraints on AGN physics as well.

In most cases, line of sight intercepting cold gas intercept more than a single cold cloud. This indicates that the total covering fraction of the cold gas within a certain galacto-centric radius could be large. However, this remains only longitudinal information. Transverse information about the distribution of cold clouds within galactic (kpc) scales is difficult to obtain, not only because of the lack of close quasar pairs, but also because the small separation required makes the separate recording of each quasar’s spectrum difficult, due to overlapping spectral PSF. By inspecting a new C I systems I recently found in the SDSS-III, I noticed the presence of a nearby point source which we then confirmed to be the lensed image of the same quasar from Nordic Optical Telescope observations. The apparent angular separation between the two images (2 arcsec) allowed us to comfortably obtain high resolution spectroscopic data with Keck for each line of sight independently, when the physical separation at the absorbers redshift is actually only 1 kpc. Not only coincident H I and metal absorption is found along both spectra, but we also detected neutral carbon and molecular hydrogen along the second line of sight. Despite similar strengths and the main absorption located within the same 150km s^{-1} range, the two C I profiles are un-correlated. This is expected given the size of such clouds (parsec) compared to the separation of the lines of sight, but the simultaneous detection allows us to put unique constraints on the cold gas volume filling factor ([Krogager et al. 2018b](#)) and corroborate the incidence statistics that we derived from the SDSS.

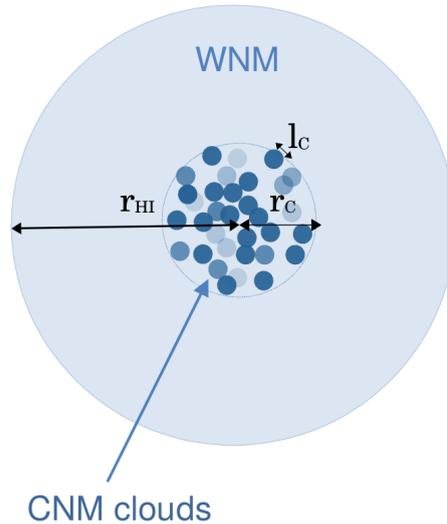


Figure 29: Illustration of the neutral gas cross-section around galaxies as can be derived from absorption observations. Lines of sight at any impact parameter within r_{HI} will feature a damped Ly- α absorption, with a column density strongly enhanced towards the centre (Noterdaeme et al. 2014). A small fraction of them ($\propto (r_C/r_{\text{HI}})^2$) also intercepts cold/molecular gas (e.g. Noterdaeme et al. 2008b; Ledoux et al. 2015; Balashev & Noterdaeme 2018). The size of each individual cold cloud (l_C) is small and so is their total volume filling factor, but their projected surface density is high so that the probability to intercept at least one of these clouds is significant within r_C . Transverse information can be obtained using multiple sight lines (Ellison et al. 2007) possibly from lensed quasars (Krogager et al. 2018b), but also using background sources with spatial extent comparable to that of the probed gas phase. Galaxies can be used to probe the overall H I (e.g. Cooke & O’Meara 2015), while the different components of quasar emission can be used to probe very small transverse scales ($l_C \sim \text{parsec}$), e.g. Balashev et al. 2011; Albornoz Vázquez et al. 2014; Klimenko et al. 2015; Bergeron & Boissé 2017). This is of course an extremely simplified picture and the actual distributions are likely highly anisotropic, following the dynamical processes that can increase the pressure locally.

With our current constraints, we can infer a picture (Fig. 29), where neutral gas is wide spread ($r_{\text{HI}} \sim \text{tens of kpc}$), and most of the absorption cross-section corresponds to warm neutral medium. Cold and molecular gas is more confined into parsec-scale clouds (l_C), and their distribution restricted within a smaller galacto-centric radius ($r_C \sim \text{a few kpc}$), where the pressure is high enough for these clouds to be formed. There is little doubt that transverse information will bring further constraints on the distribution and length-scales of the different gas phases. Future large surveys (such as the Dark Energy Spectroscopic Instrument, see DESI Collaboration 2016) should provide us with a wealth of targets to do so.

Finally, in addition to "resolving" the very small scales of cold gas clouds in the distant Universe and studying their filling factor in the larger volume, it would also be interesting to see whether we can observe time variations of the absorption lines due to relative motions of the background source, absorber and observer as suggested by Boissé et al. (2015). We can also speculate that the line of sight drift with respect to the absorber could be increased in the case of lensed systems. More important gas motions are also expected for systems associated to the quasar environment, which brings us to the discussion below about proximate systems.

Quasar environment and AGN feedback

Current models of feedback show that AGN activity can result in both quenching or triggering of star formation in galaxies. In short, quenching is expected due to gas consumption by the central engine or removal from the host galaxy by strong winds, while star-formation triggering can result from compression of the gas (e.g. Zubovas et al. 2013). However, the physics governing which outcome dominates is complex and in spite

of a wealth of observational, simulations and theoretical studies, it remains strongly debated. Absorption spectroscopy can provide unique observational constraints to understand better the properties of the gas in the quasar environment as well as the structure of the quasar itself, in particular when transverse information from partial coverage is available.

For example, we have recently identified a population of proximate damped Lyman- α absorbers that act like natural coronagraphs (Finley et al. 2013; Fathivavsari et al. 2015). In these cases, the absorber and the quasar are located at the same redshift and hence possibly physically associated as well. We are currently investigating different populations of such DLAs, with different degrees of partial coverage of the Ly- α emission, from complete absorption, partial coverage of the narrow Ly- α emission or even partial coverage of the broad emission line region (Fathivavsari et al. 2016, 2017, 2018). The Ly- α emission can then be used to map the gas from very large distances (due to easy scattering of Ly- α photons) down to the regions close to the central engine. In these latter cases, we can expect neutral gas clouds to be dense and highly excited. We indeed found signatures of excited fine-structure levels already in the SDSS spectra, which we plan to investigate at higher resolution.

In addition to neutral gas, it would be very interesting to search for molecular gas in absorption close to the quasar since H₂ is a very sensitive probe of the prevailing physical conditions. Understanding how H₂ clouds can survive or form in harsh environments could give hints on how can star-formation proceed close to the AGN. It will be interesting to understand if the H₂ gas can be formed in-situ or is carried out within flows from/to the central engine. For example, Richings & Faucher-Giguère (2018) have recently proposed that gas compression within outflows could lead to the production of H₂. More generally, the proximity of the AGN can have complex effects on the presence of molecular gas such as an increase of the dust temperature that can make H₂ form less efficiently on the surface of grains. However, the fragmentation of dust due to strong UV radiation increases the ratio of grain surface to the total dust mass hence increasing the H₂ formation. It is therefore crucial to obtain as much information as possible about the physical and chemical properties of such molecular clouds, as well as their location and kinematics with respect to the central engine to understand the formation and survival of H₂ in quasar galaxy hosts.

To do this, one should start with a systematic search for H₂ absorption systems at the quasar redshift. Since we already know that the associated damped Lyman- α line may be hidden due to leaking Ly- α emission, I propose to directly search for strong H₂ lines in low-resolution spectra. This is possible using rank-correlation analysis to recognise the typical comb-like signature of H₂. Applying this to the SDSS DR14, and comparing with a control sample of the same quasars searched by the same procedure, but shifted bluewards by 5000 km s⁻¹, we obtain the distribution of strong H₂ systems shown in Figure 30, as a function of their relative velocity to the quasar. Clearly, one can already see the neat excess of H₂ systems close to the quasar (orange histogram), compared with what would be expected from purely intervening statistics. In addition to the presence of strong H₂ lines, the SDSS data already show these have peculiar properties with very instructive AGN/absorber configurations, with for example, the Ly- α emission region from the quasar not being fully covered by the absorbing cloud, leaving a leaking Ly- α emission in the core of the DLA (see Fig. 31). In other cases, the H₂ redshift appears to be larger than that of the quasar, which could be the signature of gas infalling towards the AGN.

This may therefore just be the start of new avenue of research and following-up a sample of proximate molecular absorbers with large telescopes should provide unique, original information about the location, kinematics, physical and chemical state of cold and molecular gas close to the quasar.

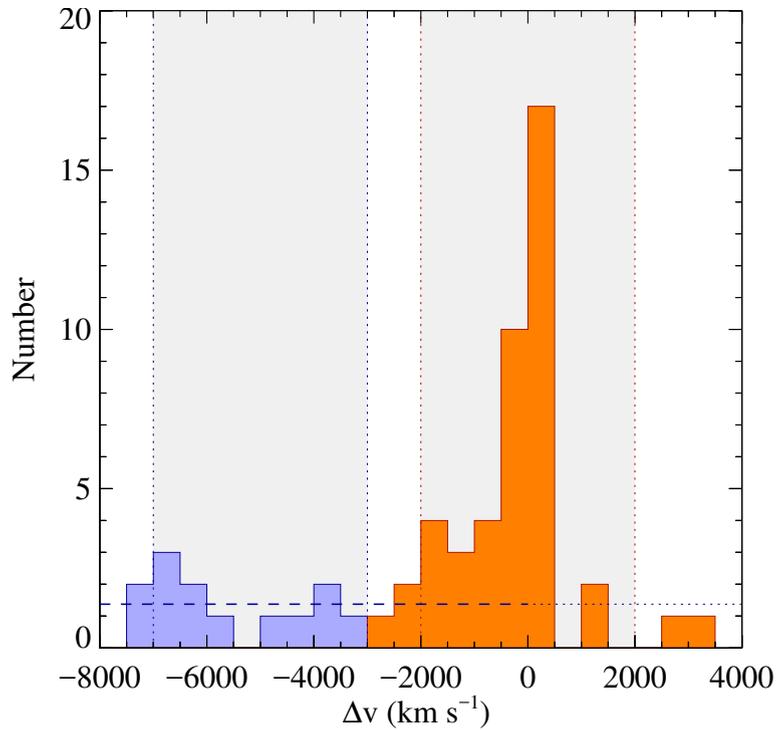


Figure 30: Number of strong damped H_2 systems as a function of their velocity with respect to the quasar systemic redshift. The search has been performed with exactly the same code applied on the same spectra, but in two different windows, illustrated by the shaded regions. Note the strong excess of H_2 systems in the central 1000 km s^{-1} compared to what would be expected if following intervening statistics (dashed line). [From Noterdaeme et al., in prep].

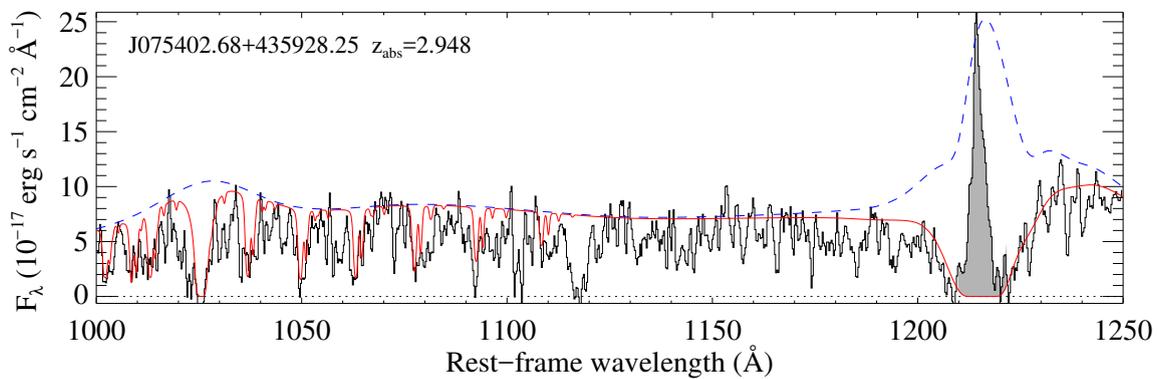


Figure 31: Strong proximate H_2 system detected in the SDSS. The red line shows the synthetic $\text{H}_2+\text{H I}$ absorption profile. The core of the damped Lyman- α absorption is filled with leaking emission from the quasar itself (shaded grey).

References

- Abgrall, H., Le Bourlot, J., Pineau Des Forets, G., Roueff, E., Flower, D. R., & Heck, L. 1992, *A&A*, 253, 525
- Albornoz Vásquez, D., Rahmani, H., Noterdaeme, P., Petitjean, P., Srianand, R., & Ledoux, C. 2014, *A&A*, 562, A88
- Altay, G., Theuns, T., Schaye, J., Booth, C. M., & Dalla Vecchia, C. 2013, *MNRAS*, 436, 2689
- Angulo, R. E., Springel, V., White, S. D. M., Jenkins, A., Baugh, C. M., & Frenk, C. S. 2012, *MNRAS*, 426, 2046
- Bahcall, J. N. & Wolf, R. A. 1968, *ApJ*, 152, 701
- Balashev, S. A., Ivanchik, A. V., & Varshalovich, D. A. 2010, *Astronomy Letters*, 36, 761
- Balashev, S. A., Klimenko, V. V., Ivanchik, A. V., Varshalovich, D. A., Petitjean, P., & Noterdaeme, P. 2014, *MNRAS*, 440, 225
- Balashev, S. A. & Noterdaeme, P. 2018, *MNRAS*, 478, L7
- Balashev, S. A., Noterdaeme, P., Klimenko, V. V., Petitjean, P., Srianand, R., Ledoux, C., Ivanchik, A. V., & Varshalovich, D. A. 2015, *A&A*, 575, L8
- Balashev, S. A., Noterdaeme, P., Rahmani, H., Klimenko, V. V., Ledoux, C., Petitjean, P., Srianand, R., Ivanchik, A. V., & Varshalovich, D. A. 2017, *MNRAS*, 470, 2890
- Balashev, S. A., Petitjean, P., Ivanchik, A. V., Ledoux, C., Srianand, R., Noterdaeme, P., & Varshalovich, D. A. 2011, *MNRAS*, 418, 357
- Bassett, B. A. & Kunz, M. 2004, *Phys. Rev. D*, 69, 101305
- Bechtold, J. & Ellingson, E. 1992, *ApJ*, 396, 20
- Bergeron, J. 1986, *A&A*, 155, L8
- Bergeron, J. & Boissé, P. 1991, *A&A*, 243, 344
- Bergeron, J. & Boissé, P. 2017, *A&A*, 604, A37
- Bialy, S., Bihl, S., Beuther, H., Henning, T., & Sternberg, A. 2017, *ApJ*, 835, 126
- Bigiel, F., Leroy, A., Walter, F., Brinks, E., de Blok, W. J. G., Madore, B., & Thornley, M. D. 2008, *AJ*, 136, 2846
- Bird, S., Vogelsberger, M., Haehnelt, M., Sijacki, D., Genel, S., Torrey, P., Springel, V., & Hernquist, L. 2014, *MNRAS*, 445, 2313
- Black, J. H., Chaffee, F. H., & Foltz, C. B. 1987, *ApJ*, 317, 442
- Boissé, P., Bergeron, J., Prochaska, J. X., Péroux, C., & York, D. G. 2015, *A&A*, 581, A109
- Boissé, P., Le Brun, V., Bergeron, J., & Deharveng, J.-M. 1998, *A&A*, 333, 841
- Boissé, P., Le Petit, F., Rollinde, E., Roueff, E., Pineau des Forêts, G., Andersson, B.-G., Gry, C., & Felenbok, P. 2005, *A&A*, 429, 509
- Bolatto, A. D., Wolfire, M., & Leroy, A. K. 2013, *ARA&A*, 51, 207
- Bouché, N., Dekel, A., Genzel, R., Genel, S., Cresci, G., Förster Schreiber, N. M., Shapiro, K. L., Davies, R. I., & Tacconi, L. 2010, *ApJ*, 718, 1001
- Bouché, N., Murphy, M. T., Kacprzak, G. G., Péroux, C., Contini, T., Martin, C. L., & Dessauges-Zavadsky, M. 2013, *Science*, 341, 50
- Bouché, N., Murphy, M. T., Péroux, C., Contini, T., Martin, C. L., Förster Schreiber, N. M., Genzel, R., Lutz, D., Gillessen, S., Tacconi, L., Davies, R., & Eisenhauer, F. 2012, *MNRAS*, 419, 2
- Bouret, J.-C., Neiner, C., Gómez de Castro, A. I., Evans, C., Gaensicke, B., Shore, S., Fossati, L., Gry, C., Charlot, S., Marin, F., Noterdaeme, P., & Chaufray, J.-Y. 2018, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Vol. 10699, *Space Telescopes and Instrumentation 2018: Ultraviolet*

- to Gamma Ray, 106993B
- Bowen, D. V., Blades, J. C., & Pettini, M. 1995, *ApJ*, 448, 634
- Braun, R. 2012, *ApJ*, 749, 87
- Cai, Z., Fan, X., Noterdaeme, P., Wang, R., McGreer, I., et al. 2014, *ApJ*, 793, 139
- Cardelli, J. A., Clayton, G. C., & Mathis, J. S. 1989, *ApJ*, 345, 245
- Carilli, C. L., Lane, W., de Bruyn, A. G., Braun, R., & Miley, G. K. 1996, *AJ*, 112, 1317
- Carswell, R. F., Becker, G. D., Jorgenson, R. A., Murphy, M. T., & Wolfe, A. M. 2012, *MNRAS*, 2703
- Carswell, R. F., Jorgenson, R. A., Wolfe, A. M., & Murphy, M. T. 2011, *MNRAS*, 411, 2319
- Cassata, P., Le Fèvre, O., Garilli, B., Maccagni, D., Le Brun, V., et al. 2011, *A&A*, 525, A143
- Catinella, B., Haynes, M. P., Giovanelli, R., Gardner, J. P., & Connolly, A. J. 2008, *ApJ*, 685, L13
- Cen, R. 2012, *ApJ*, 748, 121
- Chelouche, D. & Bowen, D. V. 2010, *ApJ*, 722, 1821
- Chen, H.-W., Prochaska, J. X., Bloom, J. S., & Thompson, I. B. 2005, *ApJ*, 634, L25
- Christensen, L., Noterdaeme, P., Petitjean, P., Ledoux, C., & Fynbo, J. P. U. 2009, *A&A*, 505, 1007
- Combes, F. 2018, *A&A Rev.*, 26, 5
- Cooke, J. & O'Meara, J. M. 2015, *ApJ*, 812, L27
- Cooke, R., Pettini, M., Steidel, C. C., Rudie, G. C., & Jorgenson, R. A. 2011, *MNRAS*, 412, 1047
- Cooke, R. J., Pettini, M., Jorgenson, R. A., Murphy, M. T., & Steidel, C. C. 2014, *ApJ*, 781, 31
- Coppolani, F., Petitjean, P., Stoehr, F., Rollinde, E., Pichon, C., Colombi, S., Haehnelt, M. G., Carswell, B., & Teysier, R. 2006, *MNRAS*, 370, 1804
- Cui, J., Bechtold, J., Ge, J., & Meyer, D. M. 2005, *ApJ*, 633, 649
- Curran, S. J. 2010, *MNRAS*, 402, 2657
- Curran, S. J., Webb, J. K., Murphy, M. T., Bandiera, R., Corbelli, E., & Flambaum, V. V. 2002, *PASA*, 19, 455
- Dalgarno, A., Black, J. H., & Weisheit, J. C. 1973, *Astrophys. Lett.*, 14, 77
- Dalgarno, A. & Stephens, T. L. 1970, *ApJ*, 160, L107
- Daprà, M., Niu, M. L., Salumbides, E. J., Murphy, M. T., & Ubachs, W. 2016, *ApJ*, 826, 192
- Daprà, M., Noterdaeme, P., Vonk, M., Murphy, M. T., & Ubachs, W. 2017, *MNRAS*, 467, 3848
- Davé, R., Finlator, K., & Oppenheimer, B. D. 2012, *MNRAS*, 421, 98
- Dawson, K. S., Kneib, J.-P., Percival, W. J., Alam, S., Albareti, F. D., et al. 2016, *AJ*, 151, 44
- Dawson, K. S., Schlegel, D. J., Ahn, C. P., Anderson, S. F., Aubourg, É., et al. 2013, *AJ*, 145, 10
- De Cia, A., Ledoux, C., Petitjean, P., & Savaglio, S. 2018, *A&A*, 611, A76
- De Cia, A., Ledoux, C., Savaglio, S., Schady, P., & Vreeswijk, P. M. 2013, *A&A*, 560, A88
- DESI Collaboration. 2016, arXiv e-prints, arXiv:1611.00036
- Dessauges-Zavadsky, M., Prochaska, J. X., D'Odorico, S., Calura, F., & Matteucci, F. 2006, *A&A*, 445, 93
- Draine, B. T. 2003, *ARA&A*, 41, 241
- Draine, B. T. & Bertoldi, F. 1996, *ApJ*, 468, 269
- Dutta, R., Srianand, R., Gupta, N., Momjian, E., Noterdaeme, P., Petitjean, P., & Rahmani, H. 2017, *MNRAS*, 465, 588
- Dutta, R., Srianand, R., Rahmani, H., Petitjean, P., Noterdaeme, P., & Ledoux, C. 2014, *MNRAS*, 440, 307
- Dvorkin, I., Vangioni, E., Silk, J., Petitjean, P., & Olive, K. A. 2016, *MNRAS*, 458, L104
- Ellison, S. L., Hall, P. B., & Lira, P. 2005, *AJ*, 130, 1345
- Ellison, S. L., Hennawi, J. F., Martin, C. L., & Sommer-Larsen, J. 2007, *MNRAS*, 378, 801
- Ellison, S. L., Yan, L., Hook, I. M., Pettini, M., Wall, J. V., & Shaver, P. 2001, *A&A*, 379, 393

- Fathivavsari, H., Petitjean, P., Jamialahmadi, N., Khosroshahi, H. G., Rahmani, H., Finley, H., Noterdaeme, P., Pâris, I., & Srianand, R. 2018, MNRAS, 477, 5625
- Fathivavsari, H., Petitjean, P., Noterdaeme, P., Pâris, I., Finley, H., López, S., & Srianand, R. 2016, MNRAS, 461, 1816
- Fathivavsari, H., Petitjean, P., Noterdaeme, P., Pâris, I., Finley, H., López, S., Srianand, R., & Sánchez, P. 2015, MNRAS, 454, 876
- Fathivavsari, H., Petitjean, P., Zou, S., Noterdaeme, P., Ledoux, C., Krühler, T., & Srianand, R. 2017, MNRAS, 466, L58
- Ferland, G. J., Porter, R. L., van Hoof, P. A. M., Williams, R. J. R., Abel, N. P., Lykins, M. L., Shaw, G., Henney, W. J., & Stancil, P. C. 2013, Rev. Mexicana Astron. Astrofis., 49, 137
- Fernández, X., Gim, H. B., van Gorkom, J. H., Yun, M. S., Momjian, E., et al. 2016, ApJ, 824, L1
- Field, G. B., Goldsmith, D. W., & Habing, H. J. 1969, ApJ, 155, L149
- Finley, H., Petitjean, P., Noterdaeme, P., & Pâris, I. 2014, A&A, 572, A31
- Finley, H., Petitjean, P., Pâris, I., Noterdaeme, P., Brinkmann, J., et al. 2013, A&A, 558, A111
- Fitzpatrick, E. L. & Massa, D. 2007, ApJ, 663, 320
- Fixsen, D. J. 2009, ApJ, 707, 916
- Flower, D. R. & Watt, G. D. 1984, MNRAS, 209, 25
- Fox, A. J., Ledoux, C., Petitjean, P., & Srianand, R. 2007, A&A, 473, 791
- Frank, S. & Péroux, C. 2010, MNRAS, 406, 2235
- Freudling, W., Staveley-Smith, L., Catinella, B., Minchin, R., Calabretta, M., Momjian, E., Zwaan, M., Meyer, M., & O’Neil, K. 2011, ApJ, 727, 40
- Fukugita, M. & Ménard, B. 2015, ApJ, 799, 195
- Fumagalli, M., O’Meara, J. M., Prochaska, J. X., Rafelski, M., & Kanekar, N. 2015, MNRAS, 446, 3178
- Fynbo, J. P. U., Geier, S. J., Christensen, L., Gallazzi, A., Krogager, J.-K., Krühler, T., Ledoux, C., Maund, J. R., Møller, P., Noterdaeme, P., Rivera-Thorsen, T., & Vestergaard, M. 2013, MNRAS, 436, 361
- Fynbo, J. P. U., Heintz, K. E., Neeleman, M., Christensen, L., Dessauges-Zavadsky, M., Kanekar, N., Møller, P., Prochaska, J. X., Rhodin, N. H. P., & Zwaan, M. 2018, MNRAS
- Fynbo, J. P. U., Jakobsson, P., Prochaska, J. X., Malesani, D., Ledoux, C., et al. 2009, ApJS, 185, 526
- Fynbo, J. P. U., Krogager, J.-K., Heintz, K. E., Geier, S., Møller, P., Noterdaeme, P., Christensen, L., Ledoux, C., & Jakobsson, P. 2017, A&A, 606, A13
- Fynbo, J. P. U., Ledoux, C., Noterdaeme, P., Christensen, L., Møller, P., et al. 2011, MNRAS, 413, 2481
- Ge, J., Bechtold, J., & Black, J. H. 1997, ApJ, 474, 67
- Ge, J., Bechtold, J., & Kulkarni, V. P. 2001, ApJ, 547, L1
- Glover, S. C. O. & Clark, P. C. 2014, MNRAS, 437, 9
- Gnedin, N. Y. & Kravtsov, A. V. 2011, ApJ, 728, 88
- Gordon, K. D., Clayton, G. C., Misselt, K. A., Landolt, A. U., & Wolff, M. J. 2003, ApJ, 594, 279
- Gry, C., Boulanger, F., Nehmé, C., Pineau des Forêts, G., Habart, E., & Falgarone, E. 2002, A&A, 391, 675
- Gry, C. & Jenkins, E. B. 2017, A&A, 598, A31
- Guimarães, R., Noterdaeme, P., Petitjean, P., Ledoux, C., Srianand, R., López, S., & Rahmani, H. 2012, AJ, 143, 147
- Guimarães, R., Petitjean, P., de Carvalho, R. R., Djorgovski, S. G., Noterdaeme, P., Castro, S., Poppe, P. C. D. R., & Aghaee, A. 2009, A&A, 508, 133
- Gupta, N., Srianand, R., Baan, W., Baker, A. J., Beswick, R. J., et al. 2016, in Proceedings of MeerKAT Science: On the Pathway to the SKA. 25-27 May, 2016 Stellenbosch, South Africa (MeerKAT2016)., 14
- Gupta, N., Srianand, R., Farnes, J. S., Pidopryhora, Y., Vivek, M., Paragi, Z., Noterdaeme, P., Oosterloo, T.,

- & Petitjean, P. 2018, *MNRAS*, 476, 2432
- Gupta, N., Srianand, R., Noterdaeme, P., Petitjean, P., & Muzahid, S. 2013, *A&A*, 558, A84
- Gupta, N., Srianand, R., Petitjean, P., Bergeron, J., Noterdaeme, P., & Muzahid, S. 2012, *A&A*, 544, A21
- Gupta, N., Srianand, R., Petitjean, P., Noterdaeme, P., & Saikia, D. J. 2009, *MNRAS*, 398, 201
- Hartoog, O. E., Fynbo, J. P. U., Kaper, L., De Cia, A., & Bagdonaite, J. 2015, *MNRAS*, 447, 2738
- Hébrard, G. & Moos, H. W. 2003, *ApJ*, 599, 297
- Heintz, K. E., Fynbo, J. P. U., Ledoux, C., Jakobsson, P., Møller, P., Christensen, L., Geier, S., Krogager, J.-K., & Noterdaeme, P. 2018, *A&A*, 615, A43
- Hollenbach, D. & Salpeter, E. E. 1971, *ApJ*, 163, 155
- Howk, J. C., Wolfe, A. M., & Prochaska, J. X. 2005, *ApJ*, 622, L81
- Hurier, G., Aghanim, N., Douspis, M., & Pointecouteau, E. 2014, *A&A*, 561, A143
- Ishiyama, T. 2014, *ApJ*, 788, 27
- Ivanchik, A. V., Petitjean, P., Balashev, S. A., Srianand, R., Varshalovich, D. A., Ledoux, C., & Noterdaeme, P. 2010, *MNRAS*, 404, 1583
- Jaeckel, J. & Ringwald, A. 2010, *Annual Review of Nuclear and Particle Science*, 60, 405
- Jetzer, P., Puy, D., Signore, M., & Tortora, C. 2011, *General Relativity and Gravitation*, 43, 1083
- Jorgenson, R. A., Murphy, M. T., Thompson, R., & Carswell, R. F. 2014, *MNRAS*, 443, 2783
- Jorgenson, R. A., Wolfe, A. M., & Prochaska, J. X. 2010, *ApJ*, 722, 460
- Jorgenson, R. A., Wolfe, A. M., Prochaska, J. X., & Carswell, R. F. 2009, *ApJ*, 704, 247
- Jorgenson, R. A., Wolfe, A. M., Prochaska, J. X., Lu, L., Howk, J. C., Cooke, J., Gawiser, E., & Gelino, D. M. 2006, *ApJ*, 646, 730
- Joshi, R., Srianand, R., Noterdaeme, P., & Petitjean, P. 2017a, *MNRAS*, 465, 701
- Joshi, R., Srianand, R., Petitjean, P., & Noterdaeme, P. 2017b, *MNRAS*, 471, 1910
- Joshi, R., Srianand, R., Petitjean, P., & Noterdaeme, P. 2018, *MNRAS*, 476, 210
- Junkkarinen, V. T., Cohen, R. D., Beaver, E. A., Burbidge, E. M., Lyons, R. W., & Madejski, G. 2004, *ApJ*, 614, 658
- Jura, M. 1974a, *ApJ*, 190, L33
- Jura, M. 1974b, *ApJ*, 191, 375
- Kanekar, N. & Chengalur, J. N. 2003, *A&A*, 399, 857
- Kashikawa, N., Misawa, T., Minowa, Y., Okoshi, K., Hattori, T., Toshikawa, J., Ishikawa, S., & Onoue, M. 2014, *ApJ*, 780, 116
- Kennicutt, R. C. & Evans, N. J. 2012, *ARA&A*, 50, 531
- Kennicutt, Jr., R. C., Calzetti, D., Walter, F., Helou, G., Hollenbach, D. J., et al. 2007, *ApJ*, 671, 333
- Khare, P., vanden Berk, D., York, D. G., Lundgren, B., & Kulkarni, V. P. 2012, *MNRAS*, 419, 1028
- Klimenko, V. V., Balashev, S. A., Ivanchik, A. V., Ledoux, C., Noterdaeme, P., Petitjean, P., Srianand, R., & Varshalovich, D. A. 2015, *MNRAS*, 448, 280
- Klimenko, V. V., Balashev, S. A., Ivanchik, A. V., & Varshalovich, D. A. 2016, *Astronomy Letters*, 42, 137
- Klypin, A. A., Trujillo-Gomez, S., & Primack, J. 2011, *ApJ*, 740, 102
- Krogager, J.-K., Fynbo, J. P. U., Heintz, K. E., Geier, S., Ledoux, C., Møller, P., Noterdaeme, P., Venemans, B. P., & Vestergaard, M. 2016a, *ApJ*, 832, 49
- Krogager, J.-K., Fynbo, J. P. U., Ledoux, C., Christensen, L., Gallazzi, A., Laursen, P., Møller, P., Noterdaeme, P., Péroux, C., Pettini, M., & Vestergaard, M. 2013, *MNRAS*, 433, 3091
- Krogager, J.-K., Fynbo, J. P. U., Møller, P., Ledoux, C., Noterdaeme, P., Christensen, L., Milvang-Jensen, B., & Sparre, M. 2012, *MNRAS*, 424, L1
- Krogager, J.-K., Fynbo, J. P. U., Noterdaeme, P., Zafar, T., Møller, P., Ledoux, C., Krühler, T., & Stockton,

- A. 2016b, MNRAS, 455, 2698
- Krogager, J.-K., Geier, S., Fynbo, J. P. U., Venemans, B. P., Ledoux, C., Møller, P., Noterdaeme, P., Vestergaard, M., Kangas, T., Pursimo, T., Saturni, F. G., & Smirnova, O. 2015, ApJS, 217, 5
- Krogager, J.-K., Gupta, N., Noterdaeme, P., Ranjan, A., Fynbo, J. P. U., Srianand, R., Petitjean, P., Combes, F., & Mahabal, A. 2018a, ApJS, 235, 10
- Krogager, J.-K., Møller, P., Fynbo, J. P. U., & Noterdaeme, P. 2017, MNRAS, 469, 2959
- Krogager, J.-K., Noterdaeme, P., O’Meara, J. M., Fumagalli, M., Fynbo, J. P. U., Prochaska, J. X., Hennawi, J., Balashev, S., Courbin, F., Rafelski, M., Smette, A., & Boissé, P. 2018b, A&A, 619, A142
- Krumholz, M. R., McKee, C. F., & Tumlinson, J. 2009, ApJ, 693, 216
- Kulkarni, S. R. & Heiles, C. 1988, Neutral hydrogen and the diffuse interstellar medium, ed. K. I. Kellermann & G. L. Verschuur, 95–153
- Kulkarni, V. P. & Fall, S. M. 2002, ApJ, 580, 732
- Lagache, G., Cousin, M., & Chatzikos, M. 2018, A&A, 609, A130
- Lagos, C. D. P., Baugh, C. M., Lacey, C. G., Benson, A. J., Kim, H.-S., & Power, C. 2011, MNRAS, 418, 1649
- Lanzetta, K. M., McMahan, R. G., Wolfe, A. M., Turnshek, D. A., Hazard, C., & Lu, L. 1991, ApJS, 77, 1
- Lanzetta, K. M., Wolfe, A. M., & Turnshek, D. A. 1995, ApJ, 440, 435
- Le Bourlot, J. 2000, A&A, 360, 656
- Le Bourlot, J., Pineau des Forêts, G., & Flower, D. R. 1999, MNRAS, 305, 802
- Le Brun, V., Bergeron, J., Boisse, P., & Deharveng, J. M. 1997, A&A, 321, 733
- Le Fèvre, O., Tasca, L. A. M., Cassata, P., Garilli, B., Le Brun, V., et al. 2015, A&A, 576, A79
- Le Petit, F., Nehmé, C., Le Bourlot, J., & Roueff, E. 2006, ApJS, 164, 506
- Ledoux, C., Noterdaeme, P., Petitjean, P., & Srianand, R. 2015, A&A, 580, A8
- Ledoux, C., Petitjean, P., Fynbo, J. P. U., Møller, P., & Srianand, R. 2006a, A&A, 457, 71
- Ledoux, C., Petitjean, P., & Srianand, R. 2003, MNRAS, 346, 209
- Ledoux, C., Petitjean, P., & Srianand, R. 2006b, ApJ, 640, L25
- Ledoux, C., Srianand, R., & Petitjean, P. 2002, A&A, 392, 781
- Leroy, A. K., Walter, F., Brinks, E., Bigiel, F., de Blok, W. J. G., Madore, B., & Thornley, M. D. 2008, AJ, 136, 2782
- Levshakov, S. A., Dessauges-Zavadsky, M., D’Odorico, S., & Molaro, P. 2002, ApJ, 565, 696
- Levshakov, S. A. & Varshalovich, D. A. 1985, MNRAS, 212, 517
- Lilly, S. J., Carollo, C. M., Pipino, A., Renzini, A., & Peng, Y. 2013, ApJ, 772, 119
- Lima, J. A. S., Silva, A. I., & Viegas, S. M. 2000, MNRAS, 312, 747
- Liszt, H. S. 2015, ApJ, 799, 66
- López, G. & Chen, H.-W. 2012, MNRAS, 419, 3553
- Lopez, S., Tejos, N., Ledoux, C., Barrientos, L. F., Sharon, K., Rigby, J. R., Gladders, M. D., Bayliss, M. B., & Pessa, I. 2018, Nature, 554, 493
- Lu, L., Sargent, W. L. W., Barlow, T. A., Churchill, C. W., & Vogt, S. S. 1996, ApJS, 107, 475
- Ma, J., Brammer, G., Ge, J., Prochaska, J. X., & Lundgren, B. 2018, ApJ, 857, L12
- Ma, J., Caucal, P., Noterdaeme, P., Ge, J., Prochaska, J. X., Ji, T., Zhang, S., Rahmani, H., Jiang, P., Schneider, D. P., Lundgren, B., & Pâris, I. 2015, MNRAS, 454, 1751
- Madau, P. & Dickinson, M. 2014, ARA&A, 52, 415
- Maeder, A. 2017, ApJ, 847, 65
- Malec, A. L., Buning, R., Murphy, M. T., Milutinovic, N., Ellison, S. L., Prochaska, J. X., Kaper, L., Tumlinson, J., Carswell, R. F., & Ubachs, W. 2010, MNRAS, 403, 1541

- Mandy, M. E. & Martin, P. G. 1993, *ApJS*, 86, 199
- Mas-Ribas, L., Miralda-Escudé, J., Pérez-Ràfols, I., Arinyo-i-Prats, A., Noterdaeme, P., Petitjean, P., Schneider, D. P., York, D. G., & Ge, J. 2017, *ApJ*, 846, 4
- McKee, C. F. & Krumholz, M. R. 2010, *ApJ*, 709, 308
- McKee, C. F. & Ostriker, J. P. 1977, *ApJ*, 218, 148
- Ménard, B., Wild, V., Nestor, D., Quider, A., Zibetti, S., Rao, S., & Turnshek, D. 2011, *MNRAS*, 417, 801
- Milutinovic, N., Ellison, S. L., Prochaska, J. X., & Tumlinson, J. 2010, *MNRAS*, 408, 2071
- Molaro, P., Levshakov, S. A., Dessauges-Zavadsky, M., & D'Odorico, S. 2002, *A&A*, 381, L64
- Møller, P., Fynbo, J. P. U., & Fall, S. M. 2004, *A&A*, 422, L33
- Møller, P. & Warren, S. J. 1993, *A&A*, 270, 43
- Moomey, D., Federman, S. R., & Sheffer, Y. 2012, *ApJ*, 744, 174
- Muller, S., Beelen, A., Black, J. H., Curran, S. J., Horellou, C., Aalto, S., Combes, F., Guélin, M., & Henkel, C. 2013, *A&A*, 551, A109
- Murphy, M. T. & Bernet, M. L. 2016, *MNRAS*, 455, 1043
- Murphy, M. T. & Liske, J. 2004, *MNRAS*, 354, L31
- Neeleman, M., Kanekar, N., Prochaska, J. X., Christensen, L., Dessauges-Zavadsky, M., Fynbo, J. P. U., Møller, P., & Zwaan, M. A. 2018, *ApJ*, 856, L12
- Neeleman, M., Prochaska, J. X., & Wolfe, A. M. 2015, *ApJ*, 800, 7
- Nestor, D. B., Rao, S. M., Turnshek, D. A., & Vanden Berk, D. 2003, *ApJ*, 595, L5
- Neufeld, D. A. & Wolfire, M. G. 2009, *ApJ*, 706, 1594
- Noterdaeme, P., Krogager, J.-K., Balashev, S., Ge, J., Gupta, N., et al. 2017, *A&A*, 597, A82
- Noterdaeme, P., Laursen, P., Petitjean, P., Vergani, S. D., Maureira, M. J., Ledoux, C., Fynbo, J. P. U., López, S., & Srianand, R. 2012a, *A&A*, 540, A63
- Noterdaeme, P., Ledoux, C., Petitjean, P., Le Petit, F., Srianand, R., & Smette, A. 2007a, *A&A*, 474, 393
- Noterdaeme, P., Ledoux, C., Petitjean, P., & Srianand, R. 2008a, *A&A*, 481, 327
- Noterdaeme, P., Ledoux, C., Srianand, R., Petitjean, P., & Lopez, S. 2009a, *A&A*, 503, 765
- Noterdaeme, P., Ledoux, C., Zou, S., Petitjean, P., Srianand, R., Balashev, S., & López, S. 2018, *A&A*, 612, A58
- Noterdaeme, P., López, S., Dumont, V., Ledoux, C., Molaro, P., & Petitjean, P. 2012b, *A&A*, 542, L33
- Noterdaeme, P., Petitjean, P., Carithers, W. C., Pâris, I., Font-Ribera, A., et al. 2012c, *A&A*, 547, L1
- Noterdaeme, P., Petitjean, P., Ledoux, C., López, S., Srianand, R., & Vergani, S. D. 2010a, *A&A*, 523, A80
- Noterdaeme, P., Petitjean, P., Ledoux, C., & Srianand, R. 2009b, *A&A*, 505, 1087
- Noterdaeme, P., Petitjean, P., Ledoux, C., Srianand, R., & Ivanchik, A. 2008b, *A&A*, 491, 397
- Noterdaeme, P., Petitjean, P., Pâris, I., Cai, Z., Finley, H., Ge, J., Pieri, M. M., & York, D. G. 2014, *A&A*, 566, A24
- Noterdaeme, P., Petitjean, P., & Srianand, R. 2015a, *A&A*, 578, L5
- Noterdaeme, P., Petitjean, P., Srianand, R., Ledoux, C., & Le Petit, F. 2007b, *A&A*, 469, 425
- Noterdaeme, P., Petitjean, P., Srianand, R., Ledoux, C., & López, S. 2011, *A&A*, 526, L7+
- Noterdaeme, P., Srianand, R., & Mohan, V. 2010b, *MNRAS*, 403, 906
- Noterdaeme, P., Srianand, R., Rahmani, H., Petitjean, P., Pâris, I., Ledoux, C., Gupta, N., & López, S. 2015b, *A&A*, 577, A24
- Nozawa, T. & Fukugita, M. 2013, *ApJ*, 770, 27
- O'Meara, J. M., Burles, S., Prochaska, J. X., Prochter, G. E., Bernstein, R. A., & Burgess, K. M. 2006, *ApJ*, 649, L61

- O'Meara, J. M., Prochaska, J. X., Burles, S., Prochter, G., Bernstein, R. A., & Burgess, K. M. 2007, *ApJ*, 656, 666
- Pei, Y. C., Fall, S. M., & Bechtold, J. 1991, *ApJ*, 378, 6
- Péroux, C., Bouché, N., Kulkarni, V. P., York, D. G., & Vladilo, G. 2011, *MNRAS*, 410, 2237
- Péroux, C., Bouché, N., Kulkarni, V. P., York, D. G., & Vladilo, G. 2012, *MNRAS*, 419, 3060
- Péroux, C., Dessauges-Zavadsky, M., D'Odorico, S., Kim, T.-S., & McMahon, R. G. 2007, *MNRAS*, 382, 177
- Péroux, C., McMahon, R. G., Storrie-Lombardi, L. J., & Irwin, M. J. 2003, *MNRAS*, 346, 1103
- Petitjean, P., Ledoux, C., Noterdaeme, P., & Srianand, R. 2006, *A&A*, 456, L9
- Petitjean, P., Srianand, R., Chand, H., Ivanchik, A., Noterdaeme, P., & Gupta, N. 2009, *Space Sci. Rev.*, 148, 289
- Petitjean, P., Srianand, R., & Ledoux, C. 2002, *MNRAS*, 332, 383
- Petitjean, P., Webb, J. K., Rauch, M., Carswell, R. F., & Lanzetta, K. 1993, *MNRAS*, 262, 499
- Pontzen, A., Governato, F., Pettini, M., Booth, C. M., Stinson, G., Wadsley, J., Brooks, A., Quinn, T., & Haehnelt, M. 2008, *MNRAS*, 390, 1349
- Pontzen, A. & Pettini, M. 2009, *MNRAS*, 393, 557
- Pottasch, S. R., Wesselius, P. R., & van Duinen, R. J. 1979, *A&A*, 74, L15
- Prochaska, J. X., Gawiser, E., Wolfe, A. M., Castro, S., & Djorgovski, S. G. 2003, *ApJ*, 595, L9
- Prochaska, J. X. & Herbert-Fort, S. 2004, *PASP*, 116, 622
- Prochaska, J. X., O'Meara, J. M., & Worseck, G. 2010, *ApJ*, 718, 392
- Prochaska, J. X., Tripp, T. M., & Howk, J. C. 2005, *ApJ*, 620, L39
- Prochaska, J. X. & Wolfe, A. M. 2009, *ApJ*, 696, 1543
- Rafelski, M., Neeleman, M., Fumagalli, M., Wolfe, A. M., & Prochaska, J. X. 2014, *ApJ*, 782, L29
- Rafelski, M., Wolfe, A. M., Prochaska, J. X., Neeleman, M., & Mendez, A. J. 2012, *ApJ*, 755, 89
- Rahmani, H., Srianand, R., Gupta, N., Petitjean, P., Noterdaeme, P., & Vásquez, D. A. 2012, *MNRAS*, 3460
- Rahmani, H., Srianand, R., Noterdaeme, P., & Petitjean, P. 2010, *MNRAS*, 409, L59
- Rahmani, H., Wendt, M., Srianand, R., Noterdaeme, P., Petitjean, P., et al. 2013, *MNRAS*, 435, 861
- Rahmati, A. & Schaye, J. 2014, *MNRAS*, 438, 529
- Rahmati, A., Schaye, J., Pawlik, A. H., & Raičević, M. 2013, *MNRAS*, 431, 2261
- Ranjan, A., Noterdaeme, P., Krogager, J.-K., Petitjean, P., Balashev, S. A., Bialy, S., Srianand, R., Gupta, N., Fynbo, J. P. U., Ledoux, C., & Laursen, P. 2018, *A&A*, 618, A184
- Rao, S. M., Belfort-Mihalyi, M., Turnshek, D. A., Monier, E. M., Nestor, D. B., & Quider, A. 2011, *MNRAS*, 416, 1215
- Rao, S. M., Nestor, D. B., Turnshek, D. A., Lane, W. M., Monier, E. M., & Bergeron, J. 2003, *ApJ*, 595, 94
- Rao, S. M., Turnshek, D. A., & Nestor, D. B. 2006, *ApJ*, 636, 610
- Rauch, M., Haehnelt, M., Bunker, A., Becker, G., Marleau, F., et al. 2008, *ApJ*, 681, 856
- Reach, W. T., Koo, B.-C., & Heiles, C. 1994, *ApJ*, 429, 672
- Richings, A. J. & Faucher-Giguère, C.-A. 2018, *MNRAS*, 474, 3673
- Roy, N., Chengalur, J. N., & Srianand, R. 2006, *MNRAS*, 365, L1
- Rudie, G. C., Newman, A. B., & Murphy, M. T. 2017, *ApJ*, 843, 98
- Saro, A., Liu, J., Mohr, J. J., Aird, K. A., Ashby, M. L. N., et al. 2014, *MNRAS*, 440, 2610
- Savage, B. D., Bohlin, R. C., Drake, J. F., & Budich, W. 1977, *ApJ*, 216, 291
- Schruba, A., Bialy, S., & Sternberg, A. 2018, *ApJ*, 862, 110
- Schruba, A., Leroy, A. K., Walter, F., Bigiel, F., Brinks, E., et al. 2011, *AJ*, 142, 37

- Selsing, J., Fynbo, J. P. U., Christensen, L., & Krogager, J.-K. 2016, *A&A*, 585, A87
- Shapley, A. E. 2011, *ARA&A*, 49, 525
- Shaw, G., Rawlins, K., & Srianand, R. 2016, *MNRAS*, 459, 3234
- Shi, Y., Helou, G., Yan, L., Armus, L., Wu, Y., Papovich, C., & Stierwalt, S. 2011, *ApJ*, 733, 87
- Silva, A. I. & Viegas, S. M. 2002, *MNRAS*, 329, 135
- Smeding, A. G. & Pottasch, S. R. 1979, *A&AS*, 35, 257
- Snow, T. P. & McCall, B. J. 2006, *ARA&A*, 44, 367
- Sobolev, A. I., Ivanchik, A. V., Varshalovich, D. A., & Balashev, S. A. 2015, *Journal of Physics Conference Series*, 661, 012013
- Sonnentrucker, P., Welty, D. E., Thorburn, J. A., & York, D. G. 2007, *ApJS*, 168, 58
- Spitzer, L. 1978, *Physical processes in the interstellar medium* (New York Wiley-Interscience, 1978. 333 p.)
- Srianand, R., Gupta, N., Petitjean, P., Noterdaeme, P., & Ledoux, C. 2010, *MNRAS*, 405, 1888
- Srianand, R., Gupta, N., Petitjean, P., Noterdaeme, P., Ledoux, C., Salter, C. J., & Saikia, D. J. 2012, *MNRAS*, 421, 651
- Srianand, R., Gupta, N., Petitjean, P., Noterdaeme, P., & Saikia, D. J. 2008a, *MNRAS*, 391, L69
- Srianand, R., Gupta, N., Rahmani, H., Momjian, E., Petitjean, P., & Noterdaeme, P. 2013, *MNRAS*, 428, 2198
- Srianand, R., Hussain, T., Noterdaeme, P., Petitjean, P., Krühler, T., Japelj, J., Pâris, I., & Kashikawa, N. 2016, *MNRAS*, 460, 634
- Srianand, R., Noterdaeme, P., Ledoux, C., & Petitjean, P. 2008b, *A&A*, 482, L39
- Srianand, R., Petitjean, P., & Ledoux, C. 2000, *Nature*, 408, 931
- Srianand, R., Petitjean, P., Ledoux, C., Ferland, G., & Shaw, G. 2005, *MNRAS*, 362, 549
- Stecher, T. P. 1965, *ApJ*, 142, 1683
- Stecher, T. P. & Williams, D. A. 1967, *ApJ*, 149, L29
- Steidel, C. C., Dickinson, M., & Persson, S. E. 1994, *ApJ*, 437, L75
- Steigman, G., Romano, D., & Tosi, M. 2007, *MNRAS*, 378, 576
- Sternberg, A., Le Petit, F., Roueff, E., & Le Bourlot, J. 2014, *ApJ*, 790, 10
- Storrie-Lombardi, L. J., McMahon, R. G., & Irwin, M. J. 1996, *MNRAS*, 283, L79
- Storrie-Lombardi, L. J. & Wolfe, A. M. 2000, *ApJ*, 543, 552
- Straka, L. A., Noterdaeme, P., Srianand, R., Nutalaya, S., Kulkarni, V. P., Khare, P., Bowen, D., Bishof, M., & York, D. G. 2015, *MNRAS*, 447, 3856
- Tripp, T. M. & Bowen, D. V. 2005, in *IAU Colloq. 199: Probing Galaxies through Quasar Absorption Lines*, ed. P. Williams, C.-G. Shu, & B. Menard, 5–23
- Tumlinson, J., Malec, A. L., Carswell, R. F., Murphy, M. T., Buning, R., Milutinovic, N., Ellison, S. L., Prochaska, J. X., Jorgenson, R. A., Ubachs, W., & Wolfe, A. M. 2010, *ApJ*, 718, L156
- Tumlinson, J., Shull, J. M., Rachford, B. L., Browning, M. K., Snow, T. P., et al. 2002, *ApJ*, 566, 857
- Turnshek, D. A., Wolfe, A. M., Lanzetta, K. M., Briggs, F. H., Cohen, R. D., Foltz, C. B., Smith, H. E., & Wilkes, B. J. 1989, *ApJ*, 344, 567
- Ubachs, W., Bagdonaite, J., Salumbides, E. J., Murphy, M. T., & Kaper, L. 2016, *Reviews of Modern Physics*, 88, 021003
- Uzan, J., Aghanim, N., & Mellier, Y. 2004, *Phys. Rev. D*, 70, 083533
- Varshalovich, D. A., Ivanchik, A. V., Petitjean, P., Srianand, R., & Ledoux, C. 2001, *Astronomy Letters*, 27, 683
- Viegas, S. M. 1995, *MNRAS*, 276, 268
- Vladilo, G., Gioannini, L., Matteucci, F., & Palla, M. 2018, *ApJ*, 868, 127

- Vladilo, G., Prochaska, J. X., & Wolfe, A. M. 2008, *A&A*, 478, 701
- Wakelam, V., Bron, E., Cazaux, S., Dulieu, F., Gry, C., et al. 2017, *Molecular Astrophysics*, 9, 1
- Wang, W.-H., Kanekar, N., & Prochaska, J. X. 2015, *MNRAS*, 448, 2832
- Warren, S. J., Møller, P., Fall, S. M., & Jakobsen, P. 2001, *MNRAS*, 326, 759
- Weatherley, S. J., Warren, S. J., Møller, P., Fall, S. M., Fynbo, J. U., & Croom, S. M. 2005, *MNRAS*, 358, 985
- Welty, D. E., Xue, R., & Wong, T. 2012, *ApJ*, 745, 173
- Wild, V., Hewett, P. C., & Pettini, M. 2006, *MNRAS*, 367, 211
- Wild, V., Hewett, P. C., & Pettini, M. 2007, *MNRAS*, 374, 292
- Wolfe, A. M., Gawiser, E., & Prochaska, J. X. 2005, *ARA&A*, 43, 861
- Wolfe, A. M., Howk, J. C., Gawiser, E., Prochaska, J. X., & Lopez, S. 2004, *ApJ*, 615, 625
- Wolfe, A. M., Lanzetta, K. M., Foltz, C. B., & Chaffee, F. H. 1995, *ApJ*, 454, 698
- Wolfe, A. M., Prochaska, J. X., & Gawiser, E. 2003, *ApJ*, 593, 215
- Wolfe, A. M., Turnshek, D. A., Lanzetta, K. M., & Lu, L. 1993, *ApJ*, 404, 480
- Wolfe, A. M., Turnshek, D. A., Smith, H. E., & Cohen, R. D. 1986, *ApJS*, 61, 249
- Wolfire, M. G., Hollenbach, D., & McKee, C. F. 2010, *ApJ*, 716, 1191
- Wolfire, M. G., Hollenbach, D., McKee, C. F., Tielens, A. G. G. M., & Bakes, E. L. O. 1995, *ApJ*, 443, 152
- Wolfire, M. G., McKee, C. F., Hollenbach, D., & Tielens, A. G. G. M. 2003, *ApJ*, 587, 278
- Wong, T. & Blitz, L. 2002, *ApJ*, 569, 157
- Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., Ressler, M. E., Cutri, R. M., et al. 2010, *AJ*, 140, 1868
- York, D. G., Adelman, J., Anderson, Jr., J. E., Anderson, S. F., Annis, J., et al. 2000, *AJ*, 120, 1579
- York, D. G., Khare, P., Vanden Berk, D., Kulkarni, V. P., Crofts, A. P. S., et al. 2006, *MNRAS*, 367, 945
- Zafar, T., Møller, P., Péroux, C., Quiret, S., Fynbo, J. P. U., Ledoux, C., & Deharveng, J.-M. 2017, *MNRAS*, 465, 1613
- Zerbi, F. M., Bouchy, F., Fynbo, J., Maiolino, R., Piskunov, N., et al. 2014, in *Proc. SPIE*, Vol. 9147, Ground-based and Airborne Instrumentation for Astronomy V, 914723
- Zhang, S., Ge, J., Jiang, P., Zhou, H., Ma, J., Brandt, W. N., York, D. G., Noterdaeme, P., & Schneider, D. P. 2015, *ApJ*, 802, 92
- Zou, S., Petitjean, P., Noterdaeme, P., Ledoux, C., Krogager, J.-K., Fathivavsari, H., Srianand, R., & López, S. 2018, *A&A*, 616, A158
- Zubovas, K., Nayakshin, S., Sazonov, S., & Sunyaev, R. 2013, *MNRAS*, 431, 793
- Zwaan, M. A. & Prochaska, J. X. 2006, *ApJ*, 643, 675
- Zwaan, M. A., van der Hulst, J. M., Briggs, F. H., Verheijen, M. A. W., & Ryan-Weber, E. V. 2005, *MNRAS*, 364, 1467