UNIVERSITY OF CALIFORNIA
SANTA CRUZ

FOOD FOR STARS: THE ROLE OF HYDROGEN IN THE FORMATION AND EVOLUTION OF GALAXIES

A dissertation submitted in partial satisfaction of the
requirements for the degree of

DOCTOR OF PHILOSOPHY

in

ASTRONOMY AND ASTROPHYSICS

by

Michele Fumagalli

September 2012

The Dissertation of Michele Fumagalli
is approved:

Professor J. Xavier Prochaska, Chair

Professor Mark R. Krumholz

Professor Arthur M. Wolfe

Professor Joel R. Primack

Tyrus Miller
Vice Provost and Dean of Graduate Studies

Table of Contents

List of Figures

List of Tables

Abstract

Food for stars: the role of hydrogen in the formation and evolution of galaxies

by

Michele Fumagalli

The current cosmological model, the ΛCDM theory, describes with remarkable precision the assembly and growth of the large scale structures and of the dark matter halos in our Universe. A comprehensive theory for the baryon processes that take place within dark matter halos is, instead, still the subject of active research. The three major ingredients of this theory are known: accretion of hydrogen from the intergalactic medium, star formation, and feedback mechanisms in the form of galactic winds. However, the recipe to blend them together has not yet been found. This thesis focuses on the role that two of these ingredients have in the assembly and evolution of galaxies. The underlying questions that this work aims to address are how the accretion of hydrogen onto galaxies occurs and what the conditions needed to convert this raw fuel into stars are. The instruments used for this investigation are diverse, because of the multiplicity of physical processes, spatial scales, and cosmic epochs involved in the problem. Theory, or more specifically the analysis of hydrodynamic simulations to unveil gas accretion onto high-redshift galaxies, is the starting point for this work. In the second part, spectroscopy of bright quasars is used to probe the physical properties of gas and metals around and within distant galaxies. These observations are systematically compared to model predictions. Deep optical imaging is also used to connect the star formation rates of these galaxies to the gas properties that are measured in absorption. Finally, in the third part, the relationship between hydrogen and star formation on smaller scales is investigated by means of multiwavelength observations of local galaxies. This thesis contributes to the aforementioned open questions in four ways. First, it

is shown that the accretion of gas onto galaxies as predicted by current simulations imprints characteristic signatures on the distribution of hydrogen and metals of a particular family of absorption line systems, the Lyman limit systems. Second, new spectroscopic observations that led to the discovery of gas clouds with physical properties that match predictions from simulations are presented, paving the way for establishing empirically how galaxies acquire their gas. Third, through a comparison of the hydrogen content and the star formation rates of distant galaxies, this thesis confirms how the presence of significant amounts of hydrogen is not a sufficient condition for the onset of star formation. Finally, after assessing the validity of star formation models in environments that are common to high redshift galaxies, these findings have been interpreted as inefficient star formation in regions with low gas column density and low metallicity.

This work is dedicated to

Silvia

who, with immense love,

let me cross the ocean

to follow my dream.

Acknowledgments

"This research was possible thank to the generous support of the J. Xavier Prochaska fellowship". Seriously. I am highly indebted to my PhD advisor for his tremendous generosity. In these four years, working with Xavier has been a unique experience. He gave me total freedom to get distracted by any topic that I considered fun at that moment. His name does not even appear on some of the papers I wrote during my PhD, but to think that he did not play a major role behind the scenes would be a great mistake. Regardless to the subject of the discussion, Xavier had always some time to meet and chat about a new finding that excited me or a problem on which I was stuck. "You bet" was his response to all my emails in which I was asking "Are you free to meet?". Clearly, he was too busy to add more, but never too busy to push me back. Xavier has been a great teacher. From him I learned how to write a proposal that can compete for time at highly oversubscribed facilities, how to prepare and conduct observations, how to reduce data, and how to turn them into a paper. I will miss the chalkboard.

I really enjoyed interacting with Mark Krumholz. In reply to my comment on how my 2009 paper with him and Leslie Hunt was largely based on other people's models – something that at that time I considered a weakness – Mark said "This is what observers are for!". From that day, I decided to become what I am. Still an observer, but with a strong emphasis on theory. A good choice after all, so thank you, Mark.

In these years, I also had the pleasure to exchange ideas and initiate collaborations with more than fifteen bright minds all around the world. A highly incomplete list includes Avishai Dekel, Art Wolfe, Joe Hennawi, John O'Meara, Joel Primack, and my "blazar friends". I would like to thank them for paying attention to what I had to say even during my scientific infancy.

It so happens that there are 24 hours in a day, and about 35040 hours in four years.

As a grad student, I spent about two third of them together with great fellows, in particular Robert, Val, and Morgan. I had a great time working and having fun with them. I will miss the "supreme-vegetarian-nachos-for-here" and the Perk breaks. My family and my friends have been fantastic during these years. I never felt far from them, despite the 5000 miles and 9 time zones in between us. Silvia F. and Peppo deserve a special note.

Silvia, I don't have words to express my gratitude for you. I am looking forward to starting a new life together.

The text of this thesis includes reprints of the following previously published material: i) Fumagalli, M., Krumholz, M. R., Prochaska, J. X., Gavazzi, G., & Boselli, A. 2009, ApJ, 697, 1811; ii) Fumagalli, M., Krumholz, M. R., & Hunt, L. K. 2010, ApJ, 722, 919; iii) Fumagalli, M., O'Meara, J. M., Prochaska, J. X., & Kanekar, N. 2010, MNRAS, 408, 362; iv) Fumagalli, M., Prochaska, J. X., Kasen, D., et al. 2011, MNRAS, 418, 1796; v) Fumagalli, M., O'Meara, J. M., & Prochaska, J. X. 2011, Science, 334, 1245. The co-authors have contributed to various aspects of the data processing, analysis, and interpretation of the findings presented in these publications. As leading author of these publications, I had a major role in the scientific direction of these projects under the guidance of my PhD advisors J. Xavier Prochaska and Mark Krumholz.

Some of the data presented herein were obtained at the W.M. Keck Observatory, which is operated as a scientific partnership among the California Institute of Technology, the University of California and the National Aeronautics and Space Administration. The Observatory was made possible by the generous financial support of the W.M. Keck Foundation. This research has made use of the NASA/IPAC Extragalactic Database (NED) which is operated by the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration. This research has made use of the GOLD Mine Database. Part of this work would not have been possible without the Sloan Digital Sky Survey. Funding for the SDSS and SDSS-II was provided by the Alfred P. Sloan Foundation, the

Participating Institutions, the National Science Foundation, the U.S. Department of Energy, the National Aeronautics and Space Administration, the Japanese Monbukagakusho, the Max Planck Society, and the Higher Education Funding Council for England. The SDSS was managed by the Astrophysical Research Consortium for the Participating Institutions.

Chapter 1

Introduction

1.1 Building a galaxy: from hydrogen to stars

The fidelity with which the present cosmological theory, the ΛCDM model, reproduces the statistical properties of the large scale structures in our Universe and the spatial distribution of galaxies is quite remarkable. Once the proper recipe for dark matter and gravity are provided, supercomputers can simulate a universe that closely resembles the real one (e.g. Springel et al. 2005, Klypin et al. 2011).

Unfortunately, the same precision does not apply to theories of galaxy formation that involve baryon physics because the processes required to build a galaxy rely on a much higher degree of complexity. The tangled effects of gravity, thermodynamics, hydrodynamics, and the interaction between matter and radiation play a role on scales that range from a few parsecs to few hundred kiloparsecs.

In the past decades, the relative contribution and importance of each of these processes have been extensively investigated using numerical simulations of galaxy formation (e.g. Katz et al. 1992, Springel & Hernquist 2003, Schaye et al. 2010, Governato et al. 2010, Davé et al. 2011,

Guedes et al. 2011, Scannapieco et al. 2012) in synergy with analytic and semi-analytic models (e.g. Kauffmann et al. 1993, Mo et al. 1998, Somerville & Primack 1999, Croton et al. 2006, De Lucia & Blaizot 2007) and thanks to the constant comparisons with samples of observed galaxies from large surveys that provide a test bed for these models.

Although there are still major challenges in creating a galaxy with *ab initio* calculations, these efforts have yielded a distilled view of the three basic steps required to form a $\gtrsim 10^{10}\mathrm{M}_{\odot}$ galaxy (e.g. Bouché et al. 2010, Davé et al. 2012). First, the most abundant form of baryonic matter, hydrogen, has to funnel inside the dark matter halo and settle in a disk. Second, this primordial fuel has to be ignited to form stars. Third, a source of positive feedback that prevents the rapid collapse of gas and favors the return of processed material outside the galactic disk should be provided.

These three steps are governed by well-known laws of physics. Nevertheless, each of them is the subject of a very active field of research. The major challenge is to understand how the physics of gas and radiation on parsec scales affects the behavior of baryons on the much larger galactic scales. Moreover, the holy grail of studies of galaxy formation is not only the possibility to model each of these steps individually, but to paste them together in a self consistent theory.

This thesis offers a contribution towards deriving a complete theory of galaxy formation. In this work, we will focus on two of the three key ingredients: the accretion of gas and the formation of stars. In the first part, we will design an observational test for the current models of gas accretion onto galaxies at $2 \lesssim z \lesssim 3$ and we will present proof-of-concept observations of this experiment. This redshift range is selected both for observational convenience (see Section 1.2.2) and because it encompasses the peak of the cosmic star formation history and thus the most active era of galaxy evolution. In the second part, we will more closely investigate the conversion of gas into stars in high redshift galaxies by connecting, for the first time in a direct

way, the hydrogen column density and the star formation rates of $z \gtrsim 2$ galaxies. Finally, in the third part, we will provide a theoretical framework based on current models of star formation to interpret our findings.

Before diving into the analysis, we wish to provide in the following sections a general overview of current theories and observations of gas accretion and star formation.

1.2 Gas accretion onto galaxies: how and when

The accretion of hydrogen from the low density intergalactic medium (IGM) to the overdense dark matter halos is the necessary starting point for the formation of a galaxy. Further, accretion is a continuous phenomenon that must occur through the life of a galaxy across cosmic time. Without replenishment, the reservoir of fuel available in gaseous disks would not sustain star formation at the current rate for longer than ~ 2 Gyr in the local Universe (Bigiel et al. 2008). At higher redshift ($z \sim 1-2$), where typical star formation rates (SFRs) are higher by a factor of 10 or 100, gas would be depleted in less than 1 Gyr and, in the most extreme cases, in less than 300 Myr (Genzel et al. 2010, Daddi et al. 2010). However, the analysis of the star formation histories (e.g. Madau et al. 1998) or of the chemical enrichment (e.g. Matteucci & Francois 1989) of galaxies require star formation to occur on longer time scales.

Despite the relevance of gas accretion in the life of a galaxy, at present we have only a limited view on how gas reaches the innermost star forming regions of a galaxy, particularly at high redshifts. Our current understanding of this phenomenon is therefore largely based on models. A review of the present status of observations and theory of accretion in distant galaxies is the subject of the next sections.

1.2.1 From hot to cold: a new paradigm for gas accretion

In the classical theory of galaxy formation (Rees & Ostriker 1977, Silk 1977), gas that is streaming towards the center of a collapsing halo is shock heated to the mass-dependent virial temperature (Rees & Ostriker 1977, Kereš et al. 2005)

$$T_{\rm vir} \sim \frac{GMm_p}{kr} \sim 10^6 \left(\frac{v_{\rm circ}}{167 \text{ km s}^{-1}}\right)^2 \text{ K}\,. \tag{1.1}$$

The fate of this hot gas is then driven by two competing effects: radiative cooling that proceeds on a density dependent timescale $t_{\rm cool}$ and the gravitational collapse that proceeds on the free-fall time $t_{\rm grav}$. In the inner and denser regions where $t_{\rm cool} < t_{\rm grav}$, the shock-heated material radiatively cools, the gas looses pressure support and settles into a disk. Star formation can thus proceed. This mode of accretion, termed "hot mode", has been at the basis of all the analytic models for galaxy formation until very recently (e.g. Kauffmann et al. 1993, Mo et al. 1998, Somerville & Primack 1999).

In more recent years, however, the analysis of stability in virial shocks (Birnboim & Dekel 2003, Dekel & Birnboim 2006) has highlighted the importance of a second mode of accretion, the so-called "cold mode", in which gas never reaches the virial temperature but it is accreted at $T \lesssim 3 \times 10^5$ K. Once a virial shock develops, if the gas cooling time is shorter than the compression rate in the post-shock material, the compression due to the post-shock infall cannot restore the pressure support that is lost via radiative cooling. Thus, the post-shock gas becomes unstable and cannot sustain the shock. Applied to galaxies, the analysis of the shock stability reveals that a stable virial shock does not develop in halos below a critical mass $M_{\rm crit} \sim 10^{12} \text{M}_\odot$. Further, at $z > 2$ and $M > M_{\rm crit}$, shocks can become locally unstable and gas can still be accreted in a cold mode through dense streams, termed cold streams or cold filaments.

The relevance of cold mode accretion has been also appreciated in numerical simulations (Kereš et al. 2005) which show how over 50% of the baryons are acquired as cold in galaxies with $M_{\rm halo} < 10^{11.4}$ $M_{\odot}$ and how cold accretion still occurs through cold streams even in more massive halos where stable shocks have developed. In the last few years, the phenomenon of cold accretion and the existence of cold streams have been confirmed in both smoothed particle hydrodynamics (SPH) and adaptive mash refinement (AMR) simulations (e.g. Dekel et al. 2009b, Kereš et al. 2009). Even though small-scale shocks that are not captured at the typical resolutions of current cosmological simulations may perturb these incoming flows, the predominance of the cold mode accretion over the hot mode is now a widely accepted result. Moreover, cold streams are currently considered the preferred way through which gas enters high redshift galaxies.

1.2.2 Observing gas and metals at high redshifts

Models provide us with a very distinct picture of how gas accretion occurs in the distant Universe. Extended filaments of cold ($T \lesssim 3 \times 10^5$ K) gas converge from the ending points of the cosmic web to the center of collapsing halos, where they can deliver hydrogen to the central disks. Since it is difficult to observe hydrogen in emission at high redshift, it is particularly challenging to obtain an unambiguous empirical verification of this theoretical picture.

The 21cm radiation coming from the hyperfine structure levels of the ground state of the atomic hydrogen is too faint to be detected even in the local universe at column densities $N_{\rm HI} \lesssim 10^{18} - 10^{19}{\rm cm}^{-2}$. It is therefore impossible to image directly the neutral hydrogen at cosmological distances and other diagnostics should be considered.

A powerful way to probe hydrogen at high redshifts and, potentially, cold gas accretion in distant galaxies is through Lyman$-\alpha$ in emission. The gravitational energy of gas that is infalling within the dark matter halos can be effectively dissipated by cooling radiation, i.e. the

recombination in the Lyman$-\alpha$ line following collisional excitation. The relevance of cooling radiation from collapsing halos and, in particular, galaxies that are fed by cold streams has been investigated extensively in the literature (e.g. Haiman et al. 2000, Fardal et al. 2001, Furlanetto et al. 2005, Dijkstra & Loeb 2009, Goerdt et al. 2010, Faucher-Giguère et al. 2010). Due to both the resonant nature of the Lyman$-\alpha$ line and to the fact that the cooling radiation is exponentially sensitive to the gas temperature, contrasting results have been found by different authors (see Faucher-Giguère et al. 2010). For this reason, it is still debated whether cooling radiation from cold flows can be detected above the current sensitivity limit of $\sim 10^{-18}$ erg s^{-1} cm^{-2} arcsec^{-2}.

A third powerful way to study not only hydrogen but also the gas phase metallicity is via absorption lines. This technique, which will be routinely adopted in this thesis, has been extensively used in the past, yielding a wealth of information on the physical properties of gas in the IGM, or around and within galaxies. The idea behind this technique is quite simple. Photons from a background source with intensity I_0 which cross a gas cloud are absorbed at characteristic frequencies of the neutral hydrogen and metal ion transitions. The transmitted flux I_ν is a function of the frequency dependent optical depth τ_ν

$$I_\nu = I_0 \exp(-\tau_\nu) \, . \tag{1.2}$$

Two pieces of information can be immediately derived from the analysis of absorption lines in a spectrum. First, comparing the frequency at which the absorption occurs to the laboratory frequency of the detected species, one can establish a precise redshift for the intervening gas cloud. Second, the presence and relative strengths of metal transitions inform on the chemical and physical state of the gas, given that the opacity is a function of the oscillator strength

f, the ion column density N and the line profile $\phi(\nu)$

$$\tau_\nu = N \frac{\pi e^2}{m_e c} f \phi(\nu) \, . \tag{1.3}$$

Apart from the obvious dependence on the column density, Equation (1.3) carries information on the temperature and turbulent velocity of the gas. The absorption profile is in fact described by the Voigt profile

$$\phi(\nu) = \frac{H(u,a)}{\Delta\nu_D \sqrt{\pi}} \, , \tag{1.4}$$

where $H(u,a)$ is the Voigt function

$$H(u,a) = \frac{a}{\pi} \int_{-\infty}^{+\infty} dy \frac{\exp(-y^2)}{(u-y)^2 + a^2} \, . \tag{1.5}$$

Here, $a = \gamma/(4\pi\Delta\nu_D)$, $u = (\nu - \nu_l)/\Delta\nu_D$, $y = v/b$ and $\Delta\nu_D = (\nu_0 b)/c$, $b^2 = (2kT/m_X + \zeta^2)$ is the broadening parameter given by that gas temperature and element mass, γ is the sum over the spontaneous emission coefficients, ν is the observed spectral frequency, and ζ the turbulent velocity.

The key advantage of using absorption lines to study gas at $z > 2$ is that both the hydrogen Lyman series and multiple metal ion transitions that lie in the rest-frame UV enter the optical domain of the spectrum. Because of the high values for the oscillator strengths of these transitions ($f > 1$ in some cases), absorption lines can probe hydrogen and strong metal lines to low levels of column densities, down to $\sim 10^{12}$ cm^{-2} in high signal to noise spectra. At the same time, the presence of weaker transitions ($f < 0.01$) at the observed wavelength still allows a measure of the column density at relatively high column densities.

Information on the H I column density and on the broadening parameter can be obtained by examining one or multiple transitions in the Lyman series. At high column densities

($\gtrsim 10^{19}$ cm^{-2}), typical of the so-called damped Lyman$-\alpha$ systems ($N_{\rm HI}> 2\times10^{20}$ cm^{-2}) or of the super Lyman limit systems ($10^{19} <N_{\rm HI}< 2\times10^{20}$ cm^{-2}), the natural broadening of the line profile imprints a characteristic wing-like shape on the strongest Lyman transitions. By fitting a Voigt function to the observed line profile, one can robustly establish the hydrogen column density of the intervening systems. At intermediate column densities, typical of the Lyman limit systems ($10^{17.2} <N_{\rm HI}< 10^{19}$ cm^{-2}), the line profile is exclusively dominated by the broadening parameter. In fact, since the line is saturated, there are only weak constraints on the hydrogen column density. Finally, at lower column densities that are typical of the Lyman$-\alpha$ forest ($N_{\rm HI}< 10^{17}$ cm^{-2}), the profiles of the high-order lines in the Lyman series are no more saturated and bear information on both the column density and the broadening parameter.

Further, the study of metal lines, combined to a measure of the hydrogen column density, yields information on the chemical composition of the gas. Besides the line profile fitting method that is used for hydrogen, two other techniques are commonly adopted in measurements of ion column densities. The first one is known as the curve-of-growth (COG) analysis (see Spitzer 1978). This method consists in deriving the equivalent widths W of the absorption lines from the continuum normalized spectra

$$W(\lambda_0) = \frac{\lambda_0}{c}\int_{-\infty}^{+\infty}\left[1-\bar{I}(v)\right]dv\,. \tag{1.6}$$

Combining Equation (1.3) and (1.6) and comparing the equivalent widths of multiple transitions with different oscillator strengths for a single ion, one can derive information on the column density and broadening parameter. The ion column density can be derived from the "linear portion" of the COG (i.e. when the lines are not saturated) or in the "square-root" section of the COG (i.e. when damping wings are present). In the saturated portion of the COG, instead, the broadening parameter and the column density are degenerate and additional information on

the line broadening is required to infer the column density.

Throughout this work, we will use a third method that is commonly adopted to measure the ion column densities, known as the apparent optical depth method (Savage & Sembach 1991). At each spectral pixel, one can convert the observed opacity into a column density through Equation (1.3). The comparison of the column density profiles of multiple lines with different strengths is an effective way to detect the presence of saturated pixels. Provided that there are no saturated pixels, the integral of the column density profiles corresponds to the ion column density in resolved lines.

1.2.3 The circumgalactic medium of star forming galaxies

In this section, we provide a brief review of the most recent observations in emission and absorption of the circumgalactic medium (CGM), i.e. the gaseous interface between the central disk and the IGM, in $z \gtrsim 2$ galaxies.

According to model predictions, cold streams arise from the cosmic web outside the virial radius and reach the innermost halo regions (the inner $\sim 15 - 20$ kpc). Gas accretion is therefore a phenomenon that occurs on large scales, even at hundreds of kiloparsecs from massive galaxies. For this reason, experiments that want to probe the existence and the properties of the gas filaments are required to map the GCM up to a few virial radii. As discussed in the previous section, the task of mapping hydrogen in distant galaxies at $z \gtrsim 2$ is a non-trivial one. Besides the fact that the large cosmological distance reduces the sensitivity and spatial resolution that experiments can achieve, galactic outflows exist in star forming galaxies on similar scales. This adds ambiguity to the interpretation of the observational results.

With deep narrow band imaging, many surveys have now uncovered a population of Lyman$-\alpha$ emitters (LAEs) at $z \gtrsim 2$ (e.g. Nilsson et al. 2009, Ciardullo et al. 2012). These are galaxies with prominent emission lines with rest-frame equivalent widths $W(\mathrm{Ly}\alpha) > 20$Å.

However, less than a quarter of the Lyman break galaxies (LBGs), i.e. optically selected galaxies, can be classified as LAEs in the redshift interval $2 \lesssim z \lesssim 3$ and this population may not be representative of a typical star-forming galaxy (e.g. Nilsson et al. 2009).

More relevant to our study are narrow band observations in the Lyman$-\alpha$ line that have uncovered the existence of extended nebular emission above the sensitivity limit of $\sim 10^{-18}$ erg s^{-1} cm^{-2} arcsec^{-2} from regions with sizes $\gtrsim 100$ kpc (Steidel et al. 2000, Matsuda et al. 2004, Matsuda et al. 2011). These Lyman$-\alpha$ blobs (LABs) exhibit a diverse set of morphologies, ranging from mostly spherically symmetric to more filamentary ones (Matsuda et al. 2011). Although the cooling radiation due to gas accretion may power the observed emission in LABs (e.g. Goerdt et al. 2010), the contribution form stellar photoionization or harder sources (such as a active galactic nuclei; AGN) may be responsible for more than half of the observed luminosity (e.g. Hayes et al. 2011, Kasen et al. 2012).

By stacking 92 optically-selected galaxies with $z \sim 2.6$, Steidel et al. (2011) have detected low-surface brightness ($10^{-18} - 10^{-17}$ erg s^{-1} cm^{-2} arcsec^{-2}) Lyman$-\alpha$ radiation from extended regions ($\sim 20 - 80$ kpc). These authors have interpreted the observed emission profiles as powered by Lyman$-\alpha$ radiation from centrally-concentrated photoionized gas that is scattered in an accelerating wind. However, radiative transfer effects complicate the interpretation of these observations that can be better described either by clumpy biconical outflows that are decelerating outward (Dijkstra & Hultman Kramer 2012) or by a combination of scattered radiation from centrally-located star formation and cooling radiation from infalling material (Kasen et al. 2012).

The study of absorption line systems (ALSs) offers a complementary technique to the imaging of the Lyman$-\alpha$ line in emission. This technique has significant advantages, but also limitations. As discussed in the previous section, a sensitive measure of both the hydrogen and metal column densities is attainable in absorption. Further, studies of ALSs are not affected

by radiative transfer effects in the same way as imaging observations of the Lyman-α line are. With ALSs one can also obtain a direct measure of the gas velocity projected along the line of sight. However, the use of ALSs yields only a local measure without any information on the spatial extent of the intervening gas. Further, this experiment can be performed only along the lines of sight to bright background sources.

The latter condition constitutes an important downside. Quasars are rare and the UV continuum of galaxies is often too faint to yield high signal-to-noise spectra that can be used to study the CGM around individual galaxies at $z > 2$. For this reason, most of our current knowledge of the typical hydrogen and metal content of the high-redshift CGM comes from the stacked spectrum of ~ 500 LBG pairs at $z \sim 2.3$ (Steidel et al. 2010). In "down-the-barrel" experiment, i.e. using the galaxy continuum to map the foreground CGM, one can see that the interstellar absorption lines are blueshifted by $\sim 100 - 200$ km s^{-1}. This indicates that the bulk of the CGM along the line of sight is outflowing. Further, pairs of background/foreground galaxies at different impact parameters show significant absorption in the transverse direction to $\gtrsim 50$ kpc for the strongest metal transitions and beyond 100 kpc for hydrogen. Combining the kinematic information along the line of sight to the high covering factor inferred in the transverse direction, Steidel et al. (2010) have concluded that the majority of the circumgalactic gas of star forming galaxies is outflowing. Further, they conclude that there is no evidence in support of the predicted mode of cold accretion. These observations have stimulated a debate in the literature of the recent years about the existence and observability of cold streams. This thesis takes part to this discussion, which we defer to Chapter 2.

Using a sample of 15 high-resolution spectra of bright quasars in fields with > 600 galaxies with spectroscopic redshifts, Rakic et al. (2011) and Rudie et al. (2012) presented a study of the hydrogen distribution in the CGM of star-forming galaxies. Using a statistical analysis, the pixel optical depth method, Rakic et al. (2011) show that the opacity in the inner

$\sim$ 100 kpc around LBGs is on order of magnitude higher than the mean opacity of a random region in the Universe. Beyond 100 kpc, this overdensity suddenly drops but remains above the background value up to 2 Mpc. Comparing the line-of-sight velocity to the Hubble flow velocity, these authors show that significant peculiar velocities exist between $20 - 300$ km s^{-1} along the line of sight. The re-analysis of the same sample by Rudie et al. (2012) using Voigt profile decomposition of the full Lyman$-\alpha$ forest yielded consistent results.

1.3 Linking gas to stars

Once the infalling gas has found its way to the center of the dark matter halo, it settles in a rotating disk whose evolution is governed by, in the simplest form, the competing effect of gravity and pressure support from turbulent motion. According to the Toomre stability criterion for axisymmetric disturbances (Toomre 1964), the disk becomes unstable and fragments, triggering star formation at high surface densities Σ, when the product $\pi G\Sigma$ exceeds the one of the epicyclic frequency and velocity dispersion, $\kappa\sigma$. Because the high cosmological inflow rates quickly rise the gas surface density (Dekel et al. 2009b), star formation is a natural consequence of accretion according to this simple criterion.

Star formation is therefore the next element we take into consideration. The physics and phenomenology of star formation is perhaps the most studied area of astrophysics. This thesis does not aim to address the problem of the onset and evolution of star formation in detail, but focuses on a rather simple, yet still open, question: what is the relation between star formation rates and hydrogen content of high redshift galaxies? Before addressing this question, we offer in the following sections a brief introduction to current observations and theories on extragalactic star formation, starting with a review of the common tracers of star forming gas.

1.3.1 Observing the star forming gas in near and distant galaxies

In the local Universe, the relationship between the SFRs and the atomic and molecular phases of the ISM can be investigated with observations in emission.

The most straightforward way to probe both the cold and warm atomic hydrogen is via 21 cm in emission. Using the appropriate quantum mechanical absorption and emission coefficients, one can derive the relation between brightness temperature $T_{\rm b}$ and column density of H I from the radiative transfer equation (Spitzer 1978, Tielens 2005):

$$N_{\rm HI} = 1.823 \times 10^{13} \int T_{\rm b}(v) \left(\frac{\tau_v}{1 - e^{-\tau}} \right) dv \, . \tag{1.7}$$

Trivially, it follows that the hydrogen column density becomes a solely function of the brightness temperature in the optically thin limit (i.e. for columns $\lesssim 10^{20}$ cm^{-2})

$$N_{\rm HI} = 1.823 \times 10^{13} \int T_{\rm b}(v) dv \, . \tag{1.8}$$

Unfortunately, it is less trivial to map the molecular gas phase, and in particular its most abundant constituent, the molecular hydrogen. Given its homonuclear nature that inhibits the formation of a permanent electric dipole moment, emission from H_2 is only permitted in warm gas (> 500 K). At low temperatures, the only possibility to directly observe molecular hydrogen is via absorption spectroscopy in the Lyman and Werner bands for H_2 (e.g. Tumlinson et al. 2002) in the UV. To obviate the difficulties of relying on space telescopes, a common surrogate for H_2 is carbon-monoxide (CO) whose rotational lines can be observed at millimeter frequencies. The ^{12}CO(J1 : 0) transition at 3 mm is commonly adopted as a tracer of H_2, assuming an empirical conversion factor between the CO line intensity and the H_2 column density, $X(CO) = N(H_2)/I(CO)$. Fortunately, different calibrations that rely on virial arguments (e.g. Larson

1981), on γ-ray emission from the interaction between cosmic rays and H_2 (e.g. Bloemen et al. 1986, Strong & Mattox 1996, Abdo et al. 2010), and on the dust column density for a given dust-to-gas ratio (e.g. Bolatto et al. 2011) yield consistent results in the Milky Way. However, an increasingly large number of observational (e.g. Leroy et al. 2011, Bolatto et al. 2011) and theoretical (e.g. Narayanan et al. 2012, Feldmann et al. 2012) studies have questioned the universality of this conversion factor in extragalactic environments. According to the latest results, $X(CO)$ needs to be corrected upward in metal poor environments where CO is more prone to photo-dissociation. Conversely, $X(CO)$ should be corrected downward in high surface density regions. For this reason, estimates of molecular masses bear significant uncertainty.

At $z > 2$, the large cosmic distance complicates experiments that are aimed to trace star forming gas. Due to the low covering factor of molecular hydrogen and because bright background sources are rare, there are only a limited number of sightlines against which molecular hydrogen has been probed in absorption (Petitjean et al. 2006, Prochaska et al. 2009a). However, the improved sensitivity of millimeter arrays has also made possible the detection of the higher rotational levels of the CO ladder in normal star forming galaxies (Tacconi et al. 2010, Daddi et al. 2010). To date, these observations provide the most direct method to measure the mass and spatial extent of the molecular gas in distant galaxies, albeit with the downsides related to the $X(CO)$ conversion factor.

Given its low emissivity, atomic hydrogen cannot be detected via 21 cm in emission much beyond $z \sim 0.1$ and one has to turn to absorption line systems. Damped Lyman$-\alpha$ (DLA) systems (Wolfe et al. 2005), absorbers with column densities $N_{\rm HI} > 2 \times 10^{20} {\rm cm}^{-2}$, are the best candidates for gas that is potentially connected to star formation. First, this gas is highly neutral (Wolfe et al. 2005) and has spin temperatures as low as ~ 100 K (Ellison et al. 2012), similar to the local neutral medium. Second, typical DLAs are enriched to $\sim 10\%$ of the solar value (Prochaska et al. 2003, Rafelski et al. 2012), orders of magnitude higher than a

typical sightline through the IGM (Schaye et al. 2003, Simcoe et al. 2004). Third, DLAs exhibit high [C II] 158 μm cooling rates as inferred from the [C II*] absorption. These values cannot be explained purely because of the photoelectric heating by the far-ultraviolet UV background (Wolfe et al. 2003), but they require an additional energy source. All these pieces of evidence hint that DLA gas is connected to star formation. One of the objectives of this thesis is to directly test this connection.

1.3.2 An empirical view of the star formation law

In this section we offer a brief overview of the observational and theoretical studies that characterize the star formation law, i.e. the relation that links the star formation activity and the gas content of galaxies. A complete review of this topic is beyond the scope of this work (see McKee & Ostriker 2007, Kennicutt & Evans 2012), and we will emphasize only the results that are relevant to our analysis.

The notion of a star formation law that links the rate with which stars form to the gas content per unit volume of a galaxy dates back to the work of Schmidt (1959; 1963) in the Milky Way. The existence of a relation between the star formation rate surface density $\Sigma_{\rm SFR}$ and the gas surface density $\Sigma_{\rm gas}$ was then established in other galaxies (Kennicutt 1998). This relation, today known as the "Schmidt" or "Kennicutt-Schmidt" (KS) law, is often expressed in the form

$$\Sigma_{\rm SFR} = A\Sigma_{\rm gas}^{n} \, . \tag{1.9}$$

Considering averaged quantities for multiple galaxies, this relation is the tightest for the total gas surface density, i.e. atomic plus molecular, and has an index $n \sim 1.4$. At face value, this law reveals that the rate per unit area with which stars form is a super-linear function of the total gas surface density. However, as discussed in Section 1.3.1, there are significant uncertainties

in the $X(CO)$ conversion factor that is needed to infer the total molecular gas mass from CO observations and these errors may contribute to both the large scatter ($\sim$ 0.3 dex) and the resulting slope. Considering other tracers of denser gas, such as the HCN(J1:0) transition, a linear relation between the star formation rate surface density and the dense molecular gas is recovered (e.g. Gao & Solomon 2004), implying a tighter link between the amount of dense molecular gas and the formation of stars.

The availability of sensitive interferometric CO and H I maps of nearby galaxies (Helfer et al. 2003, Walter et al. 2008, Leroy et al. 2009b) has allowed a more detailed investigation of the radial and sub-kpc dependence of the KS law (e.g. Wong & Blitz 2002, Boissier et al. 2003, Kennicutt et al. 2007, Fumagalli & Gavazzi 2008, Bigiel et al. 2008, Leroy et al. 2008, Schruba et al. 2010). Correlations between $\Sigma_{\rm sfr}$ and $\Sigma_{\rm gas}$ are consistently found at different scales. However, steep indexes ($1.4 - 3.1$) are typically quoted when the total gas surface density is considered, while shallower power laws ($1.0 - 1.4$) are commonly found when the molecular gas surface density is taken into account.

These differences reflect the dependence of the star formation rate surface density on the atomic and molecular gas surface density, respectively. A relationship closer to linear is found when considering Σ_{H_2}, but the current sensitivity for CO observations limits this correlation to the inner galactic regions, where the total gas column density is higher. A single power law for the molecular hydrogen may still be in place at lower column densities, as suggested by deeper stacked observations (Schruba et al. 2011). Conversely, $\Sigma_{\rm sfr}$ is only weakly correlated to $\Sigma_{\rm HI}$ (steeper index) at low total column densities, mostly because at surface densities of $\gtrsim$ 10 $M_\odot$ pc^{-2} the atomic hydrogen is converted into molecular hydrogen (see also Chapter 6 and 7). For these reasons, the observed KS law can be described by two distinct regimes: one at low and one at high column densities, with a break at $\sim$ 10 $M_\odot$ pc^{-2}. Further, if a linear slope is assumed for the KS law, the data suggest a third super-linear regime for densities

$\gtrsim 80 - 100$ $M_\odot$ pc^{-2}.

In addition to this physical reason, different recipes to correct for dust extinction, different treatments of the backgrounds, and different assumptions on the $X(CO)$ conversion factor contribute to the observed discrepancy in the power law indexes. To date, the coefficients that describe the KS law have not yet fully converged.

In star forming galaxies at $z \gtrsim 2$, the correlation between the star formation rate and molecular gas surface density for a standard $X(CO)$ conversion factor resembles the KS law for the local Universe (Daddi et al. 2010, Genzel et al. 2010), although at higher SFRs. Compact starburst and submillimeter galaxies, however, show a much higher SFR surface density for any given gas surface density, once a lower $X(CO)$ factor is assumed for galaxies that are likely in interaction. While this finding is suggestive of two distinct modes of star formation, other choices of the $X(CO)$ conversion factor (Narayanan et al. 2012) show that the data may in fact be consistent with a single universal KS law.

Recent work by Rafelski et al. (2011) has explored the KS law in the H I dominated regions of $z > 2$ star forming galaxies, under the hypothesis that the low level of star formation around LBGs arises from DLA gas. Adopting a statistical formalism to connect the star formation rate surface density in the outskirts of LBGs to the atomic hydrogen column density in DLAs, these authors find a power law relation that is consistent with the one observed in the atomic-dominated regions of local galaxies. In Chapter 5, we will attempt to connect these two quantities directly.

1.3.3 A theoretical view of the star formation law

The formulation of a theory that explains the physical origin of the KS law involves understanding a series of physical processes that occur on scales much smaller than the ones on which the observed correlations are measured. Among those, there are the balance of heating and

cooling in the ISM, the formation of molecular gas, the collapse and fragmentation of molecular clouds, and the onset of feedback that prevents clouds from rapidly collapsing under the effect of gravity. Each of these processes can be studied in detail, and here we will simply focus on the theories that capture the key features of the observed KS law (e.g. Krumholz et al. 2009b).

The first question that arises is why star formation correlates tightly with molecular hydrogen while, in fact, the two major coolants of the ISM are C II and CO. Numerical and semi-analytic models that follow the chemical and thermal evolution of the ISM (Krumholz et al. 2011, Glover & Clark 2012) reveal that H_2 is not physically associated to star formation, but rather a by-product of the balance between heating and cooling. At high column density and higher extinction, CO forms effectively and becomes the dominant coolant. At lower column densities and extinctions, CO cannot shield from the dissociating radiation and most of the cooling comes from C II. Since the formation and destruction of H_2 have the same functional dependence on density and intensity of the radiation field as the heating and cooling, the fraction of molecular hydrogen varies with extinction and density following lines of constant temperatures. Further the Bonnor-Ebert mass, i.e. the maximum mass that a gas cloud in hydrostatic equilibrium within an ambient medium can sustain, is a function of the temperature and star formation proceeds more effectively in colder regions. For these reasons, H_2 becomes a good tracer of star formation.

Having established why molecular hydrogen appears to track star forming regions, the first ingredient in developing a theoretical model for the KS law is a prescription for the molecular hydrogen formation. The problem of the formation of molecules has been investigated extensively with numerical simulations (e.g. Pelupessy et al. 2006, Gnedin et al. 2009) and analytic theory (Krumholz et al. 2008; 2009a, McKee & Krumholz 2010). These studies consistently find that the relevant condition for the formation of molecular hydrogen is the ability of the gas to shield from the dissociating Lyman Werner radiation. This introduces a dependence on the gas

column density and metallicity. Conversely, some empirical studies (e.g. Blitz & Rosolowsky 2004; 2006) have emphasized the role of hydrodynamic pressure. Further details on this debate are in Chapter 6 and 7, where we explore in more depth the problem of formation of molecules.

Using the former type of prescriptions for the molecular gas formation, Krumholz et al. (2009b) developed an analytic theory for the KS law by further requiring that only a limited fraction of gas can be converted into stars per unit of free fall time. This star formation efficiency is only a function of the internal feedback in molecular clouds (Krumholz & McKee 2005) and therefore is weakly dependent on the galaxy proprieties. Only at very high gas surface densities when the interstellar pressure overcomes the internal pressure of molecular clouds, the galaxy ISM becomes relevant in shaping the internal structure of molecular clouds.

The resulting star formation law is able to capture the basic features of the observed KS law, including the decline of star formation surface density below $\Sigma_{\rm gas} \sim 10$ M$_\odot$ pc^{-2}, a linear correlation between star formation and molecular gas surface density at intermediate $\Sigma_{\rm gas}$, and a super-linear dependence of $\Sigma_{\rm sfr}$ on $\Sigma_{\rm gas}$ above $\gtrsim 80 - 100$ M$_\odot$ pc^{-2}. More recently, to overcome some of the shortcomings of this model, other theories that emphasize the role of UV heating in regulating the thermal and vertical pressure of the ISM have been proposed (Ostriker et al. 2010).

1.4 Outline

The outline of this thesis is as it follows.

Chapter 2 In this chapter we present results from the paper "Absorption-line systems in simulated galaxies fed by cold streams" by Fumagalli et al. (2011a). Here, we address the problem of visibility of cold streams in the context of the discussion presented in Section 1.2.2. With the analysis of simulated galaxies fed by cold streams, we investigate whether absorption line

systems are a powerful technique to uncover the nature of gas accretion in the circumgalactic medium of star forming galaxies.

Chapter 3 This chapter is a reprint of the paper "Detection of pristine gas two billion years after the Big Bang" by Fumagalli et al. (2011b). In this work, we present the analysis of two Lyman limit systems at $z > 3$ with no discernible metal content. The relevance of this discovery in the context of Big Bag nucleosynthesis and the observability of cold streams is discussed.

Chapter 4 This chapter, from the paper "Directly imaging damped Lyman-α galaxies at $z > 2$ – I. Methodology and first results" by Fumagalli et al. (2010b), introduces the design and first observations of an imaging survey of quasar fields that is aimed to directly connect the gas content of damped Lyman$-\alpha$ systems to the star formation rates of their galaxy counterparts.

Chapter 5 In this chapter we present preliminary analysis of the imaging survey described in Chapter 4. We assess the *in situ* SFRs of 32 damped Lyman$-\alpha$ systems at $z \gtrsim 2$ and we discuss the connection between the absorbing gas and galaxies that are detected in emission. These results will appear in the paper "Directly imaging damped Lyman-α galaxies at $z > 2$ – II. The star formation rates of gas rich galaxies" by Fumagalli et al. (2012).

Chapter 6 This chapter reproduces the paper "Testing models for molecular gas formation in galaxies: hydrostatic pressure or gas and dust shielding?" by Fumagalli et al. (2010a) and offers the necessary theoretical framework to discuss the results of the survey presented in the previous two chapters. In particular, we focus on whether the transition between molecular hydrogen is a sensitive function of column density and metallicity, providing a test for the applicability of star formation models in distant galaxies.

Chapter 7 In this chapter we present a second test for the validity of models for the atomic to molecular hydrogen conversion. These results first appeared in "Molecular hydrogen deficiency in HI-poor galaxies and its implications for star formation" by Fumagalli et al. (2009).

Chapter 8 In this chapter, we summarize the analysis presented in this thesis and we conclude with an overview of futures prospects.

Chapter 2

Absorption-line systems in simulated galaxies fed by cold streams

This chapter presents a reprint of the paper "Absorption-line systems in simulated galaxies fed by cold stream" by Fumagalli et al. (2011a). In this work, we use hydrodynamic simulations to address the problem of visibility of gas accretion onto galaxies. The importance of cold filaments to feed forming galaxies was already recognized prior to this work (Kereš et al. 2005, Dekel & Birnboim 2006, Kereš et al. 2009). However, the fact that signatures of accretion are often unseen in the spectra of high redshift galaxies (e.g. Steidel et al. 2010) led to the conclusion that either this mode of accretion is an artifact of numerical simulations (Steidel et al. 2010) or that absorption line systems are not the adequate tool of investigation (Faucher-Giguère & Kereš 2011). Through a detailed comparison between models and observations, we show that (metal poor) Lyman limit systems largely trace cold streams. We conclude that the study of absorption line systems remains a promising way to investigate in a statistical way the mechanisms with which galaxies acquire the bulk of their gas.

2.1 Abstract

Hydro cosmological simulations reveal that massive galaxies at high redshift are fed by long narrow streams of merging galaxies and a smoother component of cold gas. We post-process seven high-resolution simulated galaxies with radiative transfer to study the absorption characteristics of the gas in galaxies and streams, in comparison with the statistics of observed absorption-line systems. We find that much of the stream gas is ionized by UV radiation from background and local stellar sources, but still optically thick ($N_{\rm HI} > 10^{17}$ cm^{-2}) so that the streams appear as Lyman-limit systems (LLSs). At $z > 3$, the fraction of neutral gas in streams becomes non-negligible, giving rise to damped Lyman-α (DLA) absorbers as well. The gas in the central and incoming galaxies remains mostly neutral, responsible for DLAs. Within one (two) virial radii, the covering factor of optically thick gas is $< 25\%$ (10%) for LLSs and $< 5\%$ (1%) for DLAs, slowly declining with time following the universal expansion. Nevertheless, galaxies and their cold streams in the studied mass range, $M_{\rm vir} = 10^{10} - 10^{12}$ M$_\odot$, account for $> 30\%$ of the observed absorbers in the foreground of quasars, the rest possibly arising from smaller galaxies or the intergalactic medium. The mean metallicity in the streams is $\sim 1\%$ solar, much lower than in the galaxies. The simulated galaxies reproduce the Lyα-absorption equivalent widths observed around Lyman-break galaxies, but they severely underpredict the equivalent widths in metal lines, suggesting that the latter may arise from outflows. We conclude that the observed metal-poor LLSs are likely detections of the predicted cold streams. Revised analysis of the observed LLSs kinematics and simulations with more massive outflows in conjunction with the inflows may enable a clearer distinction between the signatures of the various gas modes.

2.2 Introduction

Cold neutral hydrogen that fuels star formation is an important driver of galaxy evolution at all times. Recent progress in observations and simulations has remarkably improved our view of gas in the early Universe. Molecular gas in emission is detected in normal star forming galaxies to $z \sim 2.3$ (Tacconi et al. 2010, Daddi et al. 2010). Large imaging and spectroscopic surveys have increased by orders of magnitude the samples of quasars and galaxy pairs useful for probing in absorption the interstellar medium (ISM), the circumgalactic medium (CGM) and the intergalactic medium (IGM) (e.g. Adelberger et al. 2003, Hennawi et al. 2006, Prochaska & Wolfe 2009, Noterdaeme et al. 2009, Prochaska et al. 2010, Steidel et al. 2010). Similarly, the increased resolution of numerical simulations has made it possible to reproduce the morphology of galaxies especially at high redshift (e.g. Agertz et al. 2009, Dekel et al. 2009a, Ceverino et al. 2010).

Significant progress is being made on the basic issue of how gas is accreted from the IGM into galaxies. A consistent theoretical picture is emerging from both analytic theory (Birnboim & Dekel 2003, Dekel & Birnboim 2006) and numerical simulations (Kereš et al. 2005, Ocvirk et al. 2008, Kereš et al. 2009, Dekel et al. 2009b). High-redshift galaxies are predicted to acquire the bulk of their gas mass at temperatures of a few 10^4 K via long narrow streams that consist of a smooth component and small merging galaxies. However, indisputable observational evidence for this predicted dominant mode of gas input is still lacking.

Lyα emission due to cooling radiation has been proposed as a signature of cold gas accretion that powers Lyα blobs (e.g. Haiman et al. 2000, Fardal et al. 2001, Furlanetto et al. 2005, Dijkstra & Loeb 2009, Goerdt et al. 2010, Matsuda et al. 2010), but the resonant nature of the Lyα line and the different physical mechanisms that influence the total Lyα luminosity (Faucher-Giguère et al. 2010; Kasen et al., in prep.) prevent us from unambiguously connecting

the observed fluxes with the presence of cold streams. Observations of gas in absorption against background sources are, in principle, ideal to probe the CGM to uncover infalling gas. However, it has been suggested that the low covering factor of the cold filaments reduces the probability of detecting them in absorption (Faucher-Giguère & Kereš 2011, Stewart et al. 2010), especially in the presence of large scale outflows (Steidel et al. 2010).

The subject of the present paper is the problem of detecting streams of cold gas in absorption against background sources. We formulate detailed theoretical predictions of the covering factor, the cross section, the column density distribution, the kinematics, and the metallicity of the neutral gas in seven galaxies, simulated with a high-resolution AMR code in a cosmological context and analyzed in the redshift range $z = 1.4 - 4$. To facilitate the comparison between our theoretical predictions and observations, we adopt the same methodology used in observational studies of absorption line systems (ALSs) in background quasars and galaxy pairs. This work focus on the visibility of cold stream and by comparing and contrasting these predictions with observations, we aim to find the observables that are best suited to probe this mode of accretion. A companion paper (Kasen et al., in prep.) will focus on the Lyα emission properties using the same simulations.

The paper consists of three parts. In the first two sections, we briefly discuss the numerical simulations (Section 2.3) and the radiative transfer calculation used to estimate the hydrogen ionization state (Section 2.4). This part is complemented with Appendix 2.9.1, where we provide detailed comparisons of the neutral hydrogen masses and column density distributions from three different radiative transfer models. In the second part, we quantify the neutral hydrogen cross section and covering factor as functions of the gas column density and redshift (Section 2.5). In the third part, we compare more directly model predictions with observations, focusing in particular on the column density distribution and incidence of absorption line systems (Section 2.6) as well as their metallicity and kinematics (Section 2.7). A summary and

conclusions follow in Section 2.8.

Throughout this work, we assume the WMAP5 values of the cosmological parameters (Komatsu et al. 2009), consistent with the cosmology adopted in the simulations ($\Omega_{\rm m} = 0.27$, $\Omega_\Lambda = 0.73$ and $H_0 = 70$ km s^{-1} Mpc^{-1}).

2.3 The simulations

2.3.1 Simulation method

A more detailed discussion of the simulations used in this work and comparisons with properties observed in high-redshift galaxies can be found in Ceverino et al. (2010). Here, we briefly summarize the numerical procedures adopted and we review some of the uncertainties associated to these simulated galaxies that are relevant to this analysis.

The simulations have been run with the adaptive mash refinement (AMR) hydro-gravitational code ART (Kravtsov et al. 1997, Kravtsov 2003). The code incorporates many of the relevant physical processes for galaxy formation, including gas cooling and photoionization heating, star formation, metal enrichment and stellar feedback (Ceverino & Klypin 2009, Ceverino et al. 2010). Cooling rates were tabulated for a given gas density, temperature, metallicity, and UV background (UVB) based on CLOUDY (Ferland et al. 1998) and assuming cooling at the center of a one kpc thick cloud, illuminated by a uniform radiation field. Photoheating is self-consistently modeled together with radiative cooling, through the same CLOUDY tables, assuming a uniform UVB (Haardt & Madau 1996). Self-shielding of dense, galactic neutral hydrogen from the UVB is approximated by suppressing the UVB intensity to the prereionization value (5.9×10^{-26} erg s^{-1} cm^{-2} Hz^{-1} sr^{-1}) above gas densities $n_{\rm H} = 0.1$ cm^{-3}. This threshold is consistent with the results from radiative transfer calculations (see Appendix 2.9.1).

Stars form according to a stochastic model that is roughly consistent with the Kennicutt

(1998) law in cells where the gas temperature is below 10^4K and the density is above the threshold $n_{\rm H} = 1$ cm^{-3}. The code implements a feedback model, in which the combined energy from stellar winds and supernova explosions is released at a constant heating rate over 40 Myr, the typical age of the lightest star that explodes as a type II supernova (SN II). Energy injection by type Ia supernovae (SN Ia) is also included. The heating rate from SN Ia assumes an exponentially declining rate with a maximum at 1 Gyr.

The ISM is enriched by metals from SN II and SN Ia. Metals are assumed to be released from each stellar particle by SN II at a constant rate for 40 Myr since its birth, assuming a Miller & Scalo (1979) initial mass function (IMF) and matching the yields of Woosley & Weaver (1995). The metal ejection by SN Ia assumes the same exponentially declining rate. The code treats metal advection self-consistently and metals can diffuse and pollute the medium around the central disks. In these simulations, the dark matter particle mass is 5.5×10^5 M$_\odot$, the minimum star particle mass is 10^4 M$_\odot$, the smallest cell physical size is always between 35 and 70 pc.

Each dark matter halo has been selected from a cosmological dark-matter-only simulation and resimulated with the addition of baryons, using a zoom-in technique that follows the evolution of the selected lagrangian region of ~ 1 comoving Mpc^3 (for details on the initial conditions see Ceverino & Klypin 2009). All the simulated galaxies are centrals to their halos and the analysis has been performed on a box of four times the virial radius ($R_{\rm vir}$), centered at all times on the main halo. Note that the selected galaxies do not reside in a sizable group or cluster at $z = 0$ because of the difficulties in simulating companion galaxies.

Although these simulations reproduce some of the main properties observed in high-redshift galaxies, like the Tully-Fisher and the mass-SFR relations (see Ceverino et al. 2010), the use of approximated prescriptions in lieu of the complex interplay of physical processes that operate on scales smaller than our resolution limits the predictive power of these models and of simulations in general. Particularly relevant for our work is the uncertainty associated with

stellar feedback, implemented according to the prescription of Ceverino & Klypin (2009). In these simulations, outflows eject hot gas and metals at high velocity outside the central disks into the IGM and CGM (see Ceverino & Klypin 2009). The total outflow mass flux reaches, in some cases, about one-third of the inflow flux. However, these winds may not be as strong as some observations suggest (e.g. Steidel et al. 2010) and AGN feedback is missing. As a consequence, a fraction of the gas inside the disks might actually be ejected by stronger outflows, altering the covering fraction of neutral gas and the metallicity distribution in the CGM (see e.g. Faucher-Giguère et al. 2011). Unless the low density winds drastically perturb the shape and kinematics of the gas that inflows in dense narrow streams, these models capture the properties of the cold streams, but caution is advised in generalizing our results.

2.3.2 Sample properties

For each galaxy, the total dark matter and instantaneous stellar mass are computed summing all the particles within the virial radius. Similarly, we obtain the total star formation rate (SFR) within the virial radius by averaging the masses of all the stars formed in the last 60 Myr. This time interval is arbitrarily chosen to be large enough to avoid fluctuations due to the stochastic star formation rate implemented in the simulation, but the average SFR does not depend on this assumption. The total hydrogen masses are computed from the hydrogen number density in each cell of the AMR grid as $m_{\rm H} = m_{\rm p}\ n_{\rm H}\ l^3$, with $m_{\rm p}$ the proton mass and l the cell size. No correction for heavier elements, including helium, is applied throughout this work, except when we quote a gas fraction where we multiply the hydrogen mass by 1.38 to correct for heavier elements. Finally, we derive virial masses by combining together the baryonic and non-baryonic masses.

The sample is composed of seven massive galaxies that exhibit dark matter and baryonic properties roughly consistent with the observed population of Lyman break galaxies (LBGs).

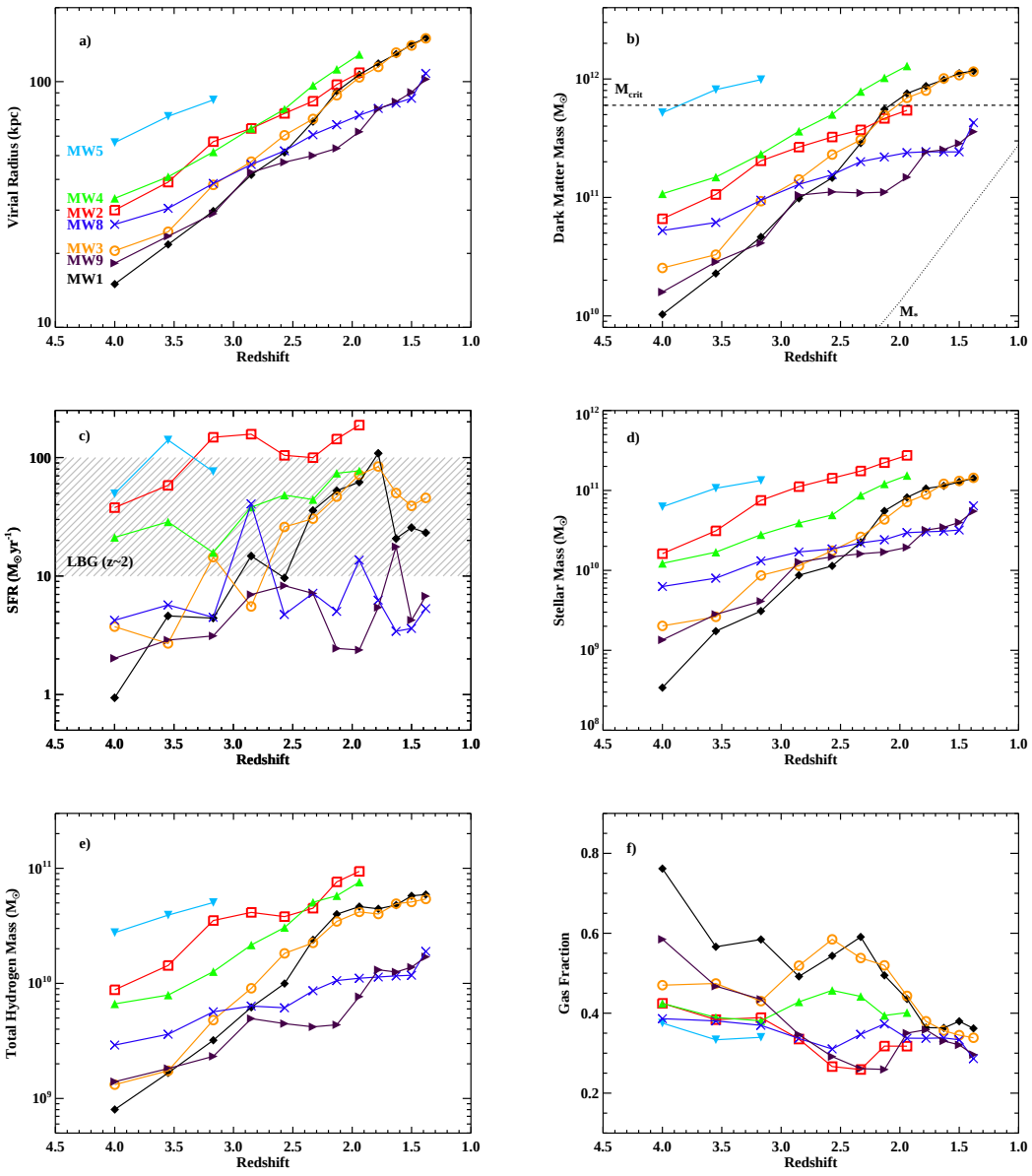

Figure 2.1: Evolution of the galaxy properties within the virial radius. Each galaxy is identified by a different line connecting different symbols. a) Virial radii. b) Dark matter mass, where the dotted line marks the characteristic halo mass, and the horizontal dashed line is the critical mass above which a stable shock can develop (Dekel & Birnboim 2006). c) Star formation rate, where the shaded area represents the typical SFR of Lyman break galaxies at $z \sim 2$ (Reddy et al. 2010) to highlight the general overlap. d) Stellar masses. e) Hydrogen masses. f) Gas fraction, defined as the ratio of gas mass to total baryon mass.

Galaxy properties as a function of time[1] are displayed in Figure 2.1 and listed in Table 2.1 and Table 2.2 in Appendix 2.9.2. Panel a) and b) show the redshift evolution of the virial radius and dark matter mass. These galaxies span roughly more than a decade in mass. As seen in panel b), they are all well above the characteristic Press-Schechter halo mass (M_*) in the relevant mass range, more so at higher redshifts. Galaxy MW1, and to some extent MW3, exhibit a faster overall growth rate than the other galaxies. The interval in mass and redshifts analyzed here is typical for cold gas accretion; only at later times ($z < 2$) halos approach the critical mass (horizontal dashed line) above which a stable shock develops and cold accretion becomes less efficient (Dekel & Birnboim 2006, Kereš et al. 2009).

The evolution of the SFR within the virial radius is presented in panel c). While there is a tendency to have higher star formation rates in higher mass halos, the stochastic nature of the simulated SFR and different accretion histories produce a less ordered star formation history (SFH). The star formation in the sample brackets the typical values found in LBGs (shaded area for $z \sim 2$; Reddy et al. 2010), with MW8 and MW9 usually below the commonly observed range. The evolution of the stellar mass is shown in panel d) and follows the general halo growth. Finally, the gas properties are displayed in panels e) and f) where we show the time evolution of the hydrogen mass and of the gas fraction, here defined as the ratio of gas mass to total baryonic mass. The simulated galaxies have a gas fraction comparable to the estimates from observed molecular gas at $z \sim 1-2$ (Tacconi et al. 2010, Daddi et al. 2010). The baryonic fraction inside the virial radius is at all times below the universal value $f_{\rm b} = 0.165$, approaching $0.25 - 0.3 f_{\rm b}$ when only gas and stars are considered inside the central disks. MW2 is an exception, having a particularly high baryon fraction at all redshifts, probably reflecting an extraordinary high gas accretion rate (Ceverino et al. 2010).

[1] Galaxies MW2, MW4, and MW5 were not available at low redshifts when this sample was assembled.

2.3.3 Satellite identification

Being part of a larger cosmological simulation, each box contains also satellites whose positions and sizes are used during the analysis to disentangle gas that resides in bound structures from gas in the streams. For this reason, we identify the satellites using the AMIGA halo finder (Gill et al. 2004, Knollmann & Knebe 2009). This code locates the centers of the halos and iteratively computes their radii so that they define an overdensity $\Delta_{\rm vir}$ over the background density $\rho_{\rm b}$. For small child halos embedded within larger parent halos, this radius is defined instead as the distance at which the density profile reaches a minimum as one moves away from the halo center. We adopt the internally computed $\Delta_{\rm vir}$ from AMIGA as a function of redshift. We consider particles with velocities greater than the escape velocity to be unbound and we select only satellites with more than 182 particles. This sets a lower limit to the satellite halo mass of $\sim 10^8$ $M_\odot$.

While AMIGA outputs several parameters for each of the identified halos, in the following we compute masses and SFRs considering all the particles that are enclosed in the defining radius, purely based on positions. Although not identical to the mass computed internally by AMIGA, these masses are well correlated with the ones generated by the code. This simple definition allows us to compute self consistently the H I masses and SFRs. Due to this approximation, our satellite masses can be uncertain up to a factor of two. Child halos contribute to the parent halo masses.

2.4 Atomic hydrogen neutral fraction

2.4.1 The radiative transfer post processing

A reliable estimate of the ionization state of the gas is essential for any study of absorption line systems. Ideally, the ionization state of the gas should be coupled to the hydrodynamic calculations, but due to the numerical complexity of this problem, codes that solve the (approximated) radiative transfer (RT) equation (e.g. Gnedin & Abel 2001, Abel & Wandelt 2002) together with the hydrodynamic equations has been developed and used only recently (e.g. Razoumov et al. 2006, Petkova & Springel 2009, Wise & Abel 2010) and they remain expensive for high resolution simulations. Currently, RT post-processing is the practical way to compute the ionization state of the gas in our high-redshift simulations. The post processing approximation is justified by the fact that the radiation-diffusion timescale is short compared to the evolution timescale.

The neutral fraction $x_{\rm HI}$ of the atomic hydrogen in each AMR cell is determined using a Monte Carlo RT code that includes both collisional ionization and photoionization due to Lyman continuum radiation from both the cosmological UVB and local stellar sources, together with absorption and scattering by dust grains. This calculation improves upon results based on simple ray-tracing schemes. In fact, Monte Carlo transport does not introduce preferred directions in the photon path and takes into account the scattering of ionizing photons. Thus, a diffuse component of the UV radiation field is naturally included (about 32% of the time, photons scatter off of hydrogen atoms rather than being absorbed). Further, radiation from stars and dust absorption is modeled locally for each stellar particle in order to reproduce the anisotropy of the radiation field within the central disks and satellites galaxies. Additional details on the numerical procedures and validating tests can be found in a companion paper (Kasen et al., in prep.). Here, we briefly summarize the models adopted for the source and sink

of ionizing radiation.

2.4.1.1 Collisional ionization

At temperatures above 10^4 K, hydrogen atoms are ionized by electron collisions. Neglecting cosmic rays that can affect the neutral fraction even at lower temperatures, electron collisions are the only factor that regulates the ionization state of the gas in the absence of external radiation fields or in regions that are optically thick to UV radiation. In our simplest model, we assume collisional ionization at equilibrium (CIE) under the simplifying assumption of a dust free gas without external radiation (Gnat & Sternberg 2007). Under this approximation, gas below $T = 10^4$ K is fully neutral, while gas above $T > 2 \times 10^4$ becomes highly ionized. In models that include also UV radiation, collisional ionization is coupled to photoionization. Since our RT calculation does not include photo-heating, the final temperature and the resulting ionization by collisions are likely to be underestimated in partially shielded regions. This problem is alleviated by the fact that the hydrodynamic code already includes a treatment for photo-heating [2].

2.4.1.2 UV radiation from cosmological background

The UV radiation from unresolved background sources is the next ingredient we add to the RT calculation. Since the UVB spectrum does not vary much with frequency over the relevant interval for hydrogen ionization, we assume a constant mean intensity $J_{912} = 5 \times 10^{-22}$ erg s^{-1} cm^{-2} Hz^{-1} sr^{-1} (Haardt & Madau 1996) over the energy range 13.6 eV $-$ 54.4 eV, corresponding to the H I ionization threshold and the cutoff at the He II threshold, respectively. Further, the modest variation as a function of redshift in the interval $z \sim 2 - 4$, within the uncertainties of the UVB models (e.g. Faucher-Giguère et al. 2008, Dall'Aglio et al. 2009),

[2] Further discussion on this issue can be found in Appendix 2.9.1.

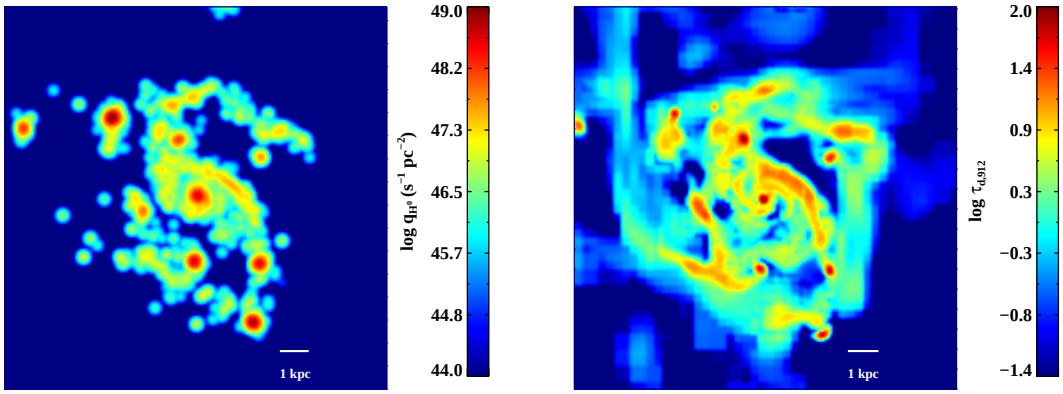

Figure 2.2: Ionizing radiation and dust in the central disk of MW3 (face on). Left panel: surface density of ionizing radiation smoothed with a Gaussian kernel of 240 pc, showing giant clumps with enhanced star formation. Right panel: integrated optical depth of dust at 912Å along the entire depth of the box (282 kpc). Star formation occurs primarily in overdense and enriched regions where the dust optical depth is high.

justifies the use of a fixed value with time. Our analysis is sensitive to the specific assumptions of the UVB, but it will become evident from the discussion that a decrease in J_{912} from $z = 3$ to $z = 4$ would not significantly affect (and even reinforce) our results.

2.4.1.3 UV radiation from local sources

An additional and important source of local ionizing radiation is the Lyman continuum from newly born stars. In our model, we estimate the rate of ionizing photons ($Q_{\rm H^0}$) emitted by a young star using stellar population synthesis models. $Q_{\rm H^0}$ is a function of the metallicity, IMF, SFH, and age of the stellar populations. Among those, the most dramatic dependence is on the SFH and age, since the ionizing radiation from a single burst can vary up to several orders of magnitude in a short time interval ($1 - 100$ Myr), when the most massive stars that contribute to the Lyman continuum luminosity leave the main sequence.

As a first approximation, we estimate the total number of ionizing photons from a constant mean SFR from the entire simulated box (i.e. from both the central galaxy and satel-

lites) using the proportionality $Q_{\mathrm{H}^0}(s^{-1}) = \alpha\ SFR(\mathrm{M}_\odot/\mathrm{yr}^{-1})$. We compute the normalization $\log\alpha = 53.168$ using STARBURTS99 (Leitherer et al. 1999), with a Kroupa IMF (Kroupa 2001) above 0.08 $\mathrm{M}_\odot$, subsolar metallicity $Z = 0.004$, and stellar models from Lejeune & Schaerer (2001). Stellar particles have different degrees of enrichment and the mean metallicity ranges from $Z = 0.004$ to $Z = 0.013$ in our sample, but the variation of Q_{H^0} with metallicity is at most a factor of two across the range $Z = 0.001 - 0.04$ ($\log\alpha = 53.227 - 52.935$), and even less ($\sim$ 0.1 dex) with the choice of the available stellar models. We redistribute the total Q_{H^0} among stellar particles which are younger than 20 Myr, in a fraction that is proportional to the stellar mass $M_{i,\mathrm{star}}$

$$Q_{i,\mathrm{H}^0} = Q_{\mathrm{H}^0}\frac{M_{i,\mathrm{star}}(t < 20\ \mathrm{Myr})}{\sum_i M_{i,\mathrm{star}}(t < 20\ \mathrm{Myr})}\,. \tag{2.1}$$

Beside the intrinsic uncertainties in the stellar models and metallicity, other reasons make the photoionization rate from stars difficult to estimate. For example, our finite resolution affects the ability to properly resolve the clumpy ISM at scales below $35 - 70$ pc. This enhances the leakage of Lyman continuum photons due to the absence of high density structures around stellar particles, but at the same time the escape fraction of ionizing photons is decreased by the lack of porosity (Wood et al. 2005). A proper estimate of the error associated with the leakage of UV photons is difficult, but most of the ionizing radiation from stellar sources is absorbed in the proximity of the emitting regions at high hydrogen volume density. In these models, the escape fraction at the virial radius is between $5 - 10$ %, varying with redshift and from galaxy to galaxy. As consequence, only a small fraction of the escaping radiation ionizes gas near the central disks with $N_{\mathrm{HI}} \sim 10^{18} - 10^{20}$ cm^{-2} (see next section). Therefore, the associated uncertainties on the neutral gas and the cumulative cross section of optically thick hydrogen beyond R_{vir} with $N_{\mathrm{HI}} \gtrsim 10^{17}$ cm^{-2} are expected to be minor.

The left panel of Figure 2.2 shows the ionizing radiation rate surface density (convolved

with a Gaussian kernel of 240 pc only for visualization purposes), for the central disk of MW3 at $z = 2.3$ and in a nearly face-on view. Most of the ionizing radiation is produced in clumps and in spiral density wakes. While the SFR in the entire box approaches 50 $M_\odot$ yr^{-1}, the central disk accounts for only half of the total ionizing radiation (with SFR ~ 27 $M_\odot$ yr^{-1}). About 50% of the remaining photons come from a large satellite which is forming stars at ~ 10 $M_\odot$ yr^{-1}. This example highlights the fact that photons from young stars that escape the surrounding halos can introduce low level anisotropies in the UV radiation field.

2.4.1.4 Dust opacity

Dust is an important sink of Lyman continuum photons, particularly relevant for radiation from local sources. Using the metallicity in the gas phase from SN II and SN Ia, the dust volume density is given by

$$\rho_{\rm d} = f_{\rm d}\ \mu\ m_{\rm p}\ (Z_{\rm SN\ Ia} + Z_{\rm SN\ II})(n_{\rm HI} + 0.01\ n_{\rm HII})\,, \tag{2.2}$$

where $f_{\rm d} = 0.4$ is an estimate of the fraction of metals locked in dust (Dwek 1998) and $\mu = 1.245$ is the mean particle weight, including helium. To mimic grain destruction, we suppress dust formation in ionized regions to only 1% of what is found in the neutral phase. This fraction is the most uncertain quantity in the dust model (see Laursen et al. 2009; section 7).

From the dust density we compute the dust optical depth at 912Å, i.e. the hydrogen ionization potential, in each AMR cell as $\tau = \alpha\ l$, with $\alpha = n_{\rm d}\ (\sigma_{\rm s} + \sigma_{\rm a}) = \kappa\ \rho_{\rm d}\ (1 - A)^{-1}$ and l the cell size. Here, $\sigma_{\rm s}$ and $\sigma_{\rm a}$ are the scattering and absorption cross sections, A is the albedo, and κ is the frequency dependent dust absorptive opacity. For the above quantities, we assume $\kappa = 9.37 \times 10^4$ cm^2 g^{-1} and $A = \sigma_{\rm s}/\sigma_{\rm d} = 0.248$ (Draine 2003). During the RT calculations, at the relevant UV wavelengths, a linear fit to the Li & Draine (2001) data is used:

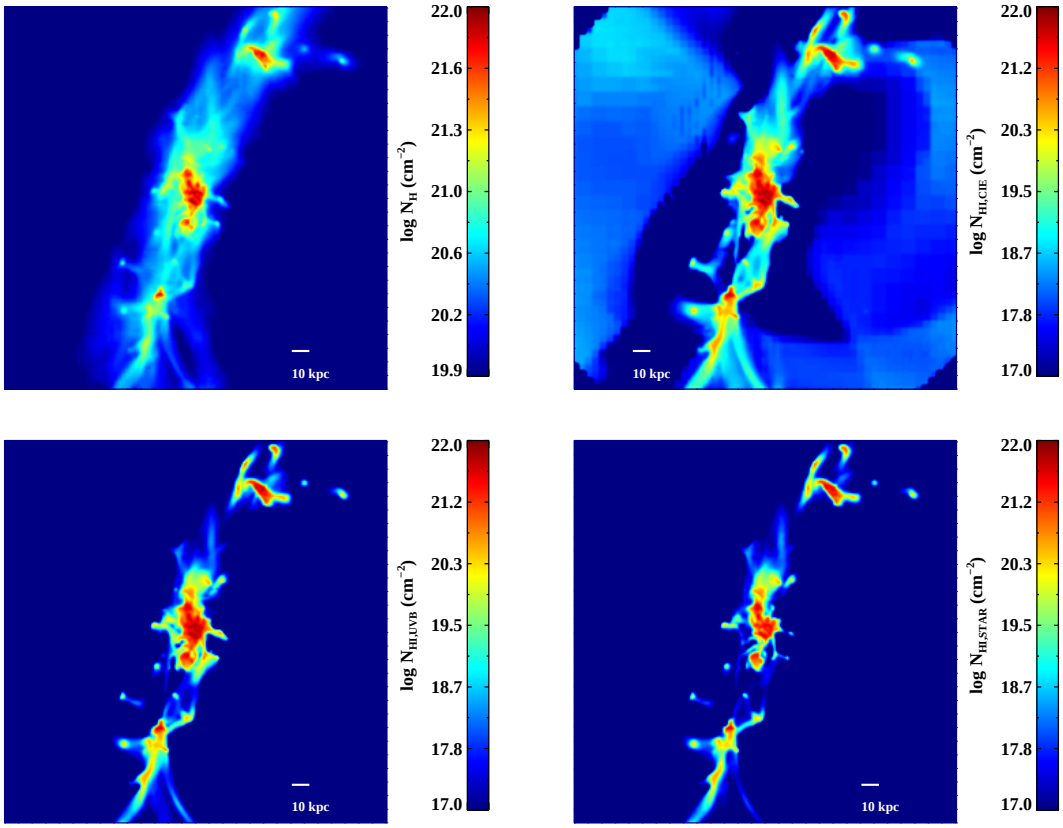

Figure 2.3: Hydrogen column density for MW3 at $z = 2.3$. Top left: $N_{\rm H}$. Top right: $N_{\rm HI}$ from the CIE model. Bottom left: $N_{\rm HI}$ from the UVB model. Bottom right: $N_{\rm HI}$ from the STAR model. Most of the gas that resides in the streams is ionized by electron collisions and the UVB, while photons from newly born stars affect the high column density inside the central and satellite galaxies and their immediate surroundings.

$\kappa(\lambda) = 9.25 \times 10^4 - 91.25 \times (912\mathring{A} - \lambda)$ and $A(\lambda) = 0.24 + 0.00028(912\mathring{A} - \lambda)$. A map of the dust optical depth for MW3 at $z = 2.3$ in a nearly face-on view is in the right panel of Figure 2.2. Here the projected dust optical depth is computed along a path of 282 kpc, the size of the entire box.

2.4.2 Results of the radiative transfer calculation

For each galaxy, we run three different RT models, gradually including additional physical processes. In the first model (hereafter CIE model), we derive the neutral fraction assuming CIE, without any source of radiation. In the second calculation (UVB model), we include the UVB together with dust and collisional ionization. Finally, in our third model (STAR model) we add ionizing radiation from local sources to the UVB model. Figure 2.3 presents an output from these calculations. In the top left panel, we show the projected $N_{\rm H}$ column density for MW3 at $z = 2.3$. High column density gas is accreting onto the central galaxy through large radial streams, with gas overdensities associated with satellites (two incoming galaxies along the streams and two closer in, near the central disk). In the other three panels, we display the $N_{\rm HI}$ column density from the different RT models.

This figure captures the basic differences that arise from the different physical processes included in the RT calculation. Part of the gas within the filaments has a temperature above $\sim 10^4$ K, and collisional ionization alone (top right panel) lowers the neutral column density by more than one order of magnitude. A comparison of the H I map for the CIE and UVB models (bottom left) clearly shows that the CIE approximation largely overestimates the neutral fraction and that photoionization from the UV background affects most of the low density gas in the streams. Indeed, the filaments are highly ionized, with patches of self-shielded neutral gas that surround the main galaxy and the satellites. Cold streams are not entirely self-shielded. Finally, the inclusion of local sources mostly affects the high column density gas (where stars form) and their immediate surroundings, where the column density is caused to drop below $N_{\rm HI} \sim 10^{20}$ cm^{-2}. The low escape fraction from the galaxy disks (below 10% at the virial radius) implies a minor effect on the gas in the streams beyond $R_{\rm vir}$ without any appreciable difference compared to the UVB model for column densities below $N_{\rm HI} \sim 10^{18}$ cm^{-2}.

A more quantitative comparison between the different models is presented in Appendix 2.9.1. There, we discuss the typical volume density for self-shielding, and the effect of local sources on the column density and mass of neutral hydrogen. We also provide a crude fitting formula to the UVB model useful to improve the CIE approximation.

2.5 The projected area of neutral gas

In the previous section, we have shown that the streams are only partially shielded and that both the UVB and local sources are necessary ingredients to study the neutral and ionized gas distribution across a large range of column density. In this section, we quantify the cross section of neutral gas in and around massive galaxies and the evolution of the covering factor. In the remaining of our analysis, we consider the STAR model as our fiducial RT calculation, providing from time-to-time a comparison with results from the UVB model.

2.5.1 Neutral hydrogen cross section

We derive the cross section that these simulations offers in projection as a function of the H I column density by collapsing the AMR 3D grid onto a 2D regular grid with cell size equal to the size of the finest level in the AMR structure. We measure in each galaxy the covered area in $N_{\rm HI}$ bins of 0.2 dex, averaged in three orthogonal directions along the axes of the AMR grid. An example for MW3 at $z = 2.3$ (STAR model[3]) is in Figure 2.4. For relative comparisons, we introduce in the right hand side axis the covering factor, defined as the cross section normalized to the area inside a circle of radius $2R_{\rm vir}$. The solid line represents the cross section measured in the entire box, while the dot-dashed line highlights the contribution of the cold streams alone, obtained by masking all the gas cells inside one quarter of the virial

[3]We highlight how the three RT calculations affect the cross section at various column densities in Figure 2.17.

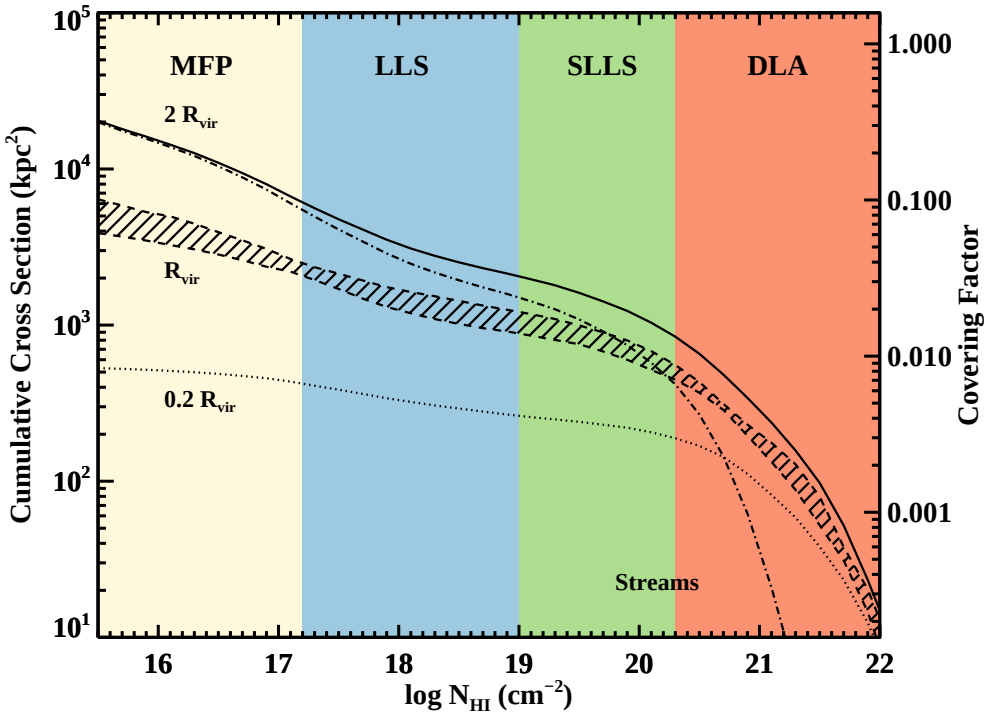

Figure 2.4: Cumulative cross section for MW3 at $z \sim 2.3$ (STAR model) averaged along three orthogonal directions. The different curves are for a cylinder of radius $2R_{\rm vir}$(solid line), a cylinder of radius $0.2R_{\rm vir}$ (dotted line) and for the streams only within $2R_{\rm vir}$ (dash dotted line). The hatched region between dashed lines highlights the dispersion along the three projection axis for a cylinder of radius $1R_{\rm vir}$. On the right axis, the physical area is translated into a fractional covering factor, normalized to the area within $2R_{\rm vir}$. Colors represent intervals of column density for DLAs, SLLSs, LLSs and MFP gas. Gas within galaxies contributes to half of the DLA cross section, while LLSs and MFP gas are mainly associated with cold streams. Only $\sim 10\%$ of the projected area is covered by optically thick gas and $\sim 1\%$ of the cross section is covered by primarily neutral gas at $N_{\rm HI} \gtrsim 10^{20}\,{\rm cm}^{-2}$).

radius of the main halos and satellites. Further, to isolate the contribution of the central galaxy, we display the cross section within two concentric cylinders of sizes $0.2R_{\rm vir}$ (dotted line) and $R_{\rm vir}$ (dashed lines). In the latter case, we display the minimum and maximum cross section among the three projection axis with a hatched region. Obviously, different orientations have different cross section distributions, but the dispersion along the mean appears not to exceed the amplitude of the features that are visible in the distribution. The radius $0.25R_{\rm vir}$ is well suited to separate streams and galaxies and it encompasses the interphase region where the flowing gas is reconnecting with the central disk. However, the fraction of neutral gas in the column density interval $10^{19} - 10^{20}$ cm^{-2} that is associated with the streams depends on the location of this boundary.

We mark in this figure four $N_{\rm HI}$ intervals that define the ionization state of the atomic hydrogen and that are associated with different classifications of observed ALSs. This is done to highlight at what intervals of column densities cold streams are predicted to dominate the neutral gas cross section. Specifically, at $N_{\rm HI} \geq 2 \times 10^{20}$ cm^{-2}, the hydrogen is neutral, highly optically thick ($\tau \gtrsim 1000$), and gives rise to damped Lyman-α absorbers (DLAs). The interval 10^{19} cm$^{-2} \leq N_{\rm HI} < 2{\times}10^{20}$ cm^{-2} marks the transition between fully neutral to ionized hydrogen and defines the super Lyman limit systems (SLLSs) or sub-damped Lyman-α absorbers. The interval 1.6×10^{17} cm$^{-2} \leq N_{\rm HI} < 10^{19}$ cm^{-2} defines Lyman limit systems (LLSs) which are highly ionized, but retain enough neutral hydrogen to be optically thick ($1 \lesssim \tau \lesssim 60$). Finally, below $N_{\rm HI} = 1.6 \times 10^{17}$ cm^{-2}, the hydrogen becomes optically thin ($\tau \lesssim 1$) and it is highly ionized. However, this gas is estimated to contribute to the mean free path (MFP) of Lyman photons in the Universe at $z > 3$ (Prochaska et al. 2009b; 2010). In this work, we consider the interval 3.2×10^{15} cm$^{-2} \leq N_{\rm HI} < 1.6 \times 10^{17}$ cm^{-2}, which we dub as "mean free path" gas.

The cross section distribution within $R_{\rm vir}$ steeply rises from the high end towards $N_{\rm HI} \sim 10^{20}$ cm^{-2} and then flattens between $N_{\rm HI} \sim 10^{18} - 10^{20}$ cm^{-2} where gas is ionized.

At lower column density there is a mild increase. Gas within $2R_{\rm vir}$ exhibits a similar shape in its cross section, but the separation between the two distributions increase systematically moving towards low column densities. From the example provided in Figure 2.4 we find that at $2R_{\rm vir}$, $\sim$ 10% of the projected area is covered by optically thick gas and $\sim$ 1% is covered by predominantly neutral gas. Note that despite the lower covering factor, this gas may still contribute to the integrated opacity of the ionizing radiation (see below). About 30% of the area is occupied by gas with $N_{\rm HI} \gtrsim 10^{15}$ cm^{-2}. Most of the MFP gas and of the LLSs are associated with the streams alone, and half of the area that is covered by SLLSs resides outside the virial halo of the massive central galaxy, with streams accounting for two third of the covering factor. A significant fraction of the neutral gas is located within halos with half of the DLA cross section inside the streams. Above $N_{\rm HI} \sim 10^{21}$ cm^{-2}, the fraction of DLAs in the streams drops below 20% and becomes negligible at even higher column densities. Only 20% of the DLA cross section is found within $0.2R_{\rm vir}$, where it is likely associated with the central disk, indicating that a significant fraction of DLAs can be found within clumps and satellites inside the virial radius of the main halo, but outside the central disk.

That a non-negligible fraction of the DLA cross section is found within the virial radius but outside the central disks appears to be a typical property of these simulations, as shown in Figure 2.5. In each snapshot (STAR model), we identify an independent structure as an ensemble of contiguous cells above $N_{\rm HI} = 2 \times 10^{20}$ cm^{-2}. We define the central galaxy to be the closest structure to the center of the main halo and we consider satellites or clumps in the streams to be all of the remaining groups of contiguous cells. The average cross section as a function of the halo mass is shown with blue squares for the central galaxies and black circles for the central galaxies and satellites. The amplitude of the halo mass intervals, chosen to have an equal number of objects, is indicated by the horizontal error bar, while the vertical error bar is for the standard deviation along the mean. Values are listed in Table 2.6.

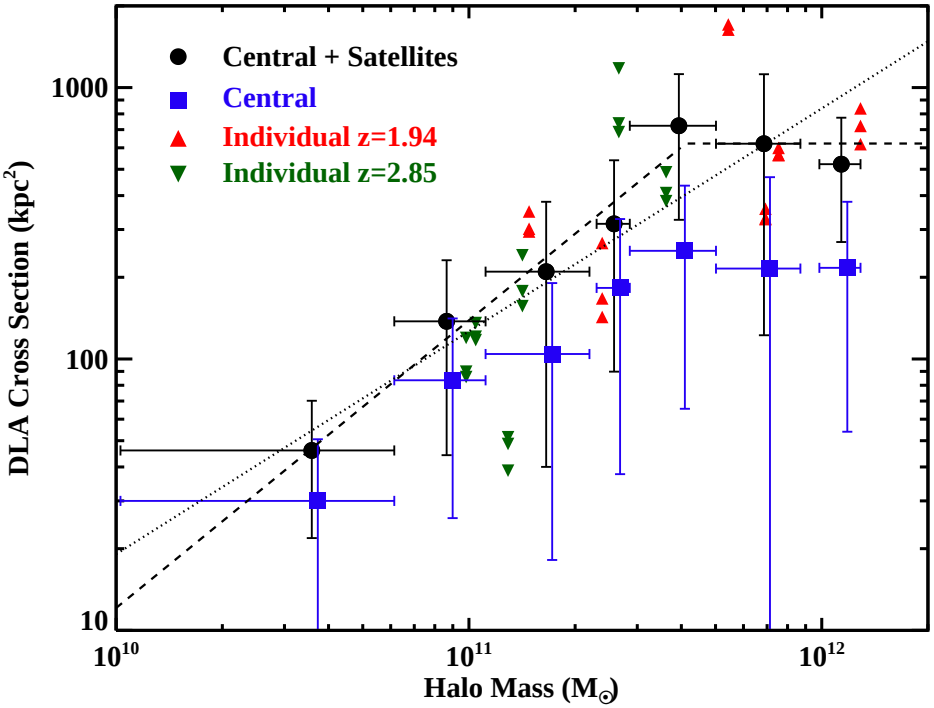

Figure 2.5: Average DLA cross section within $R_{\rm vir}$ as a function of halo mass for the central galaxies only (blue squares) and central galaxies plus satellites (black circles). All the galaxies in our sample, at all sampled redshifts, are included. The vertical error bar indicates the standard deviation about the mean, while the horizontal error bar indicates the interval of halo mass used in the average. Also shown, individual galaxies at $z \sim 2.8$ (green downward triangle) and $z \sim 1.9$ (red upward triangle). A linear regression (dashed line) shows that the total cross section is nearly proportional to the halo mass below 4×10^{11} $M_{\odot}$, while it is roughly constant at higher masses. A shallower dependence is found over the entire available mass range (dotted line), consistent with DLAs originating above a fixed surface density threshold for a self-similar gas density distribution. Central galaxies typically account for less than 50% of the DLA mean cross section.

In the entire sample, the central galaxy contributes $\sim 35\% - 65\%$ of the DLA cross section, with galaxies residing in more massive halos showing smaller fractions. We conclude that in these simulations the satellites of massive halos and clumps that reside in the streams are at least as important to the DLA cross-section of massive dark matter halos as the central galaxy (see also Maller et al. 2001, Razoumov et al. 2006). Note that the number of objects included in this study is limited and the cross-sections listed here can be affected by sample variance. Also, the cross section of DLAs is particularly sensitive to the disk sizes that are notoriously difficult to correctly reproduce in simulations, although our simulations show extended disks with $3-5$ kpc radii at high-redshift. (Ceverino et al. 2010). Further, these simulations have been selected to study cold gas accretion in massive galaxies and are not representative of the entire population of galaxies giving rise to ALSs. Nevertheless, it is reassuring that the mean DLA cross section within $R_{\rm vir}$ ranges between $50 - 800$ kpc^2 for halos of $10^{10} - 10^{11}$ M$_\odot$, in good agreement with the cross sections found at similar masses in the larger cosmological simulation by Pontzen et al. (2008) and Tescari et al. (2009). Our predictions are towards the lower end of the size distribution from Cen (2012), who finds a larger area as a consequence of strong feedback. This difference emphasis how various feedback prescriptions affect the final cross sections.

The total area is nearly proportional to the halo mass below 4×10^{11} M$_\odot$ (slope $\gamma = 1.06$ and normalization $\beta = 3.16 \times 10^{10}$ for a log-log linear regression; dashed line in Figure 2.5) and roughly constant at higher masses. Across the whole available mass range, an approximate fit to the points in Figure 2.5 is given by the mass dependence of the area within the virial radius at a fixed time: $A_{\rm vir} \propto R_{\rm vir}^2 \propto M_{\rm vir}^{2/3}$. This is consistent with DLAs originating from regions where the surface density is above a fixed threshold, provided that the distribution of densities within the virial radius (or $2R_{\rm vir}$) is self-similar among the halos.

Note that the relation found in Figure 2.5 is based on only seven independent halos at all times and redshift is partially degenerate with the halo mass, since the most massive halos are

found only at late times. It is therefore useful to inspect values for single galaxies, superimposed in this figure at $z \sim 2.8$ and $z \sim 1.9$ in three projections. While a mild evolution with redshift is visible (see also next section), the mean cross sections previously derived appear to be a fair representation of this sample as a whole, more so considering the large intrinsic scatter from galaxy to galaxy.

2.5.2 Redshift evolution of the covering factor

A qualitative description of the covering factor evolution within $2R_{\rm vir}$ is provided in Figure 2.6, where we show a gallery of projected H I column densities (STAR model) for a selected number of galaxies. Redshift decreases from the top to the bottom and mass at fixed redshift increases from the left to the right. DLAs are marked in red, SLLSs in green, LLSs in blue and MFP gas is in orange. In each panel, the galaxy center and virial radius are indicated by a dotted circle. Gas that surrounds the galaxies is non-uniform and the cold streams are "patchy" with neutral regions that are embedded in a widespread ionized medium.

An evolution in redshift appears from this figure. The gas coverage is larger at early times than at late times. Evolution in the gas concentration is evident for DLAs. At highest redshifts, there are large regions of neutral gas within the streams. These become more confined to the central disks and satellites at later times. A trend in mass is also visible, although less pronounced, with larger gas coverage in lower mass halos. Also, the topology of the streams seems to evolve with time and mass. In low mass systems and at early times, wide streams penetrate inside the virial radius to deliver cold gas to the forming galaxies. At lower redshifts and in more massive systems, the streams break into a series of narrow filaments (see Kereš & Hernquist 2009, Faucher-Giguère & Kereš 2011, Stewart et al. 2010). In our simulations, numerical resolution could be responsible for the increased smoothness in the smallest galaxies at the highest redshifts and higher resolution is required to confirm whether this trend is real.

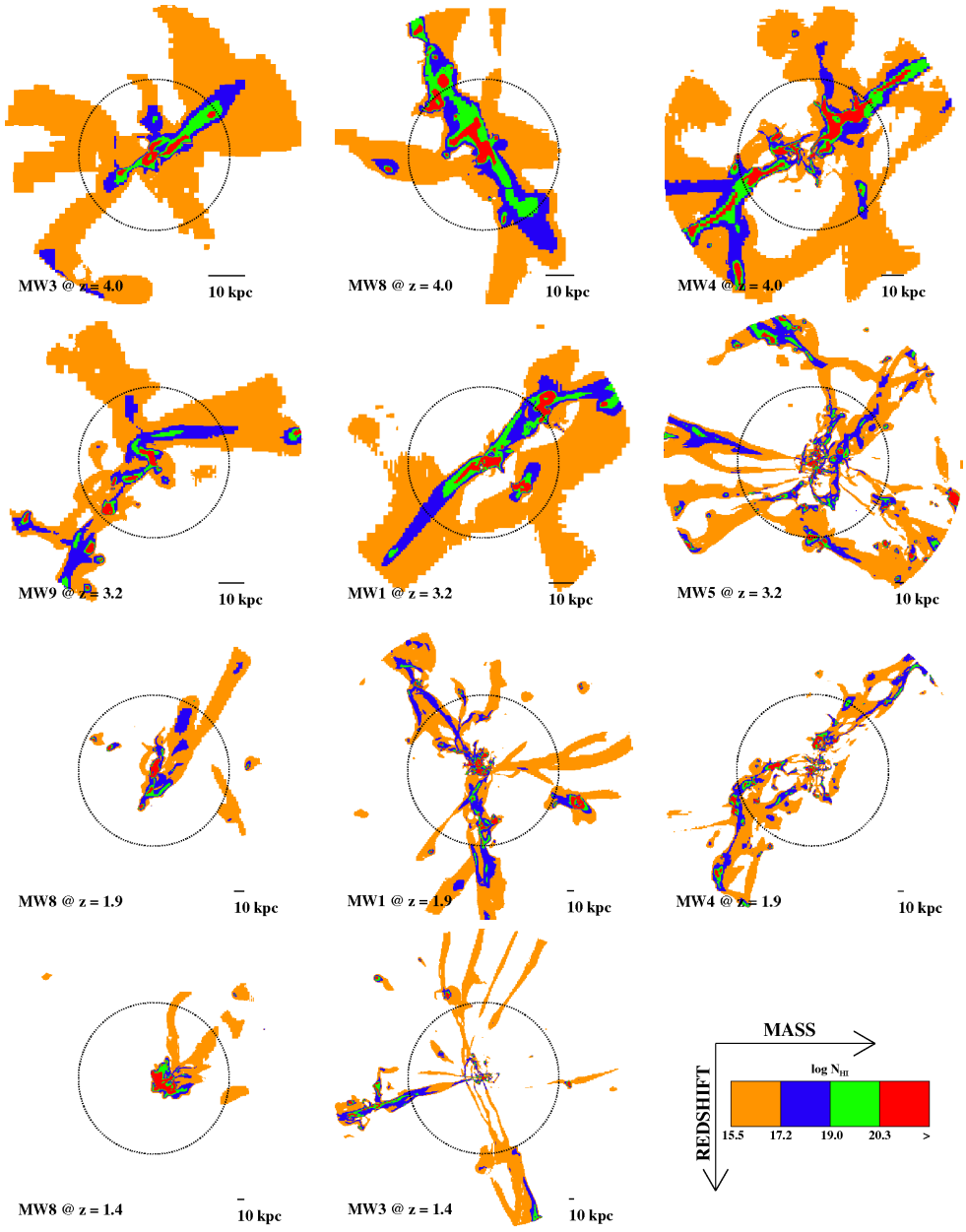

Figure 2.6: Gallery of projected H I column density (STAR model) in selected galaxies and redshifts. Four intervals of column density are marked with different colors (DLAs in red, SLLSs in green, LLSs in blue and MFP gas in orange, similar to Figure 2.4). Redshift is decreasing from top to bottom, and virial mass is increasing at constant redshift from left to right. The dotted circles mark the virial radius. Cold streams are "patchy", with neutral pockets of gas embedded in a more widespread ionized medium. A difference is seen between $z > 3$ and $z < 2$, and a marginal mass dependence is hinted.

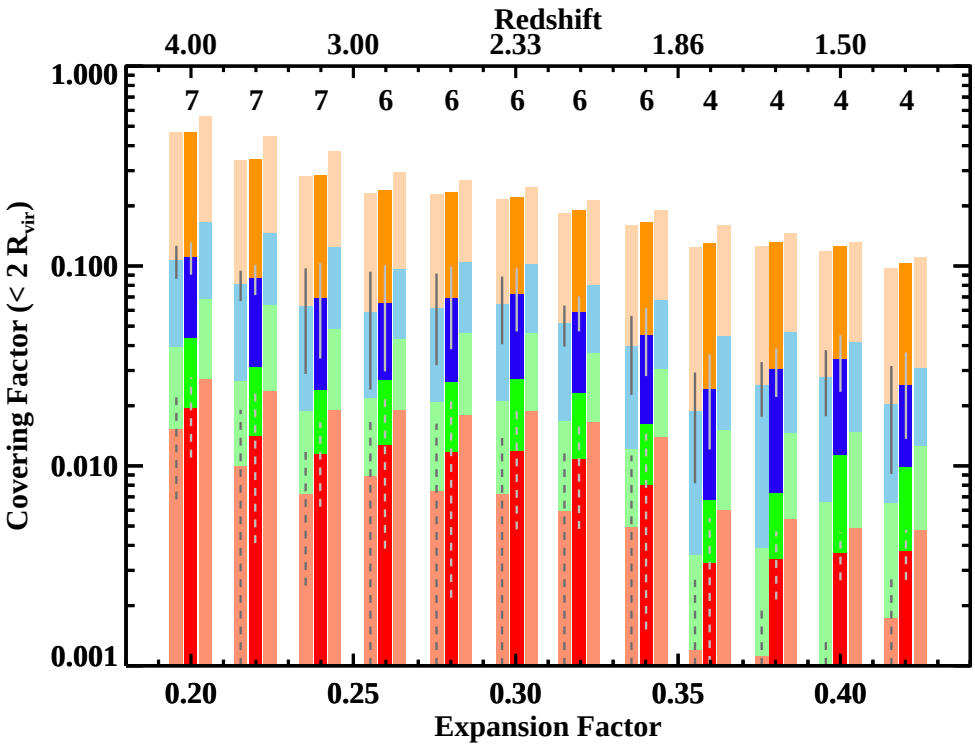

Figure 2.7: Time evolution of the cumulative covering factor within $2R_{\rm vir}$. DLAs are in red, SLLSs in green, LLSs in blue, and MFP gas in orange. The central columns are for the galaxies and streams in the STAR model, the left columns are for the streams alone in the STAR model and the right columns are for the UVB model, galaxies and streams. The numbers above the bars indicate the number of galaxies in each bin of redshift. Gray dashed and solid lines indicate the standard deviation for the mean covering factor of DLAs and LLSs respectively. The covering factor slowly decreases with redshift at all column densities, reflecting the decline of cosmological density.

A more quantitative analysis of the redshift dependence of the covering factor is given in Figure 2.7. Each bar, represents the mean values for DLAs (red), SLLSs (green), LLSs (blue) and MFP gas (orange) at one redshift and within $2R_{\rm vir}$. The central column is for the galaxies and streams in the STAR model, the left column is for the streams alone in the STAR model and the right column is for the UVB model (streams and galaxies). Values for DLAs and LLSs in individual galaxies are listed in Table 2.4. All covering factors are given as cumulative, i.e. integrated from the lowest column density to infinity (numerically approximated to $N_{\rm HI} = 10^{25}$ cm^{-2}). The numbers at the top of the bars indicate the number of galaxies included in each redshift bin. The standard deviation of the sample is shown for DLAs and LLSs with gray dashed or solid lines to highlight the wide scatter from galaxy to galaxy, partially associated to the mass (see Figure 2.9).

Figure 2.7 confirms the impression that the covering factor within $2R_{\rm vir}$ is slowly declining with time. At the highest redshifts, $30-50\%$ of the area is covered by MFP gas, $7-10\%$ is covered by LLSs and $1-2\%$ is occupied by DLAs (compare with Faucher-Giguère & Kereš 2011). By $z \sim 2$, the covering factor becomes 1% for DLAs, 5% for LLSs and $\sim 20\%$ for MFP gas. In the range $z = 2-1.3$, the covering factors are $\sim 10\%$, $3-4\%$, and $0.3-0.4\%$ respectively. Note that the already limited statistics are dominated by lower mass galaxies below $z \sim 1.9$, while the highest redshift bins may be more prone to resolution effects.

In the redshift interval $2 < z < 3$, where the sample is more homogeneous, the time evolution of the covering factor can be parametrized by $\sim a^{-d}$, with a the expansion factor. This trend can be explained by the universal expansion for a self-similar density distribution of neutral hydrogen. In these simulations, the cumulative H I column density PDF can be described by $\propto N_{\rm HI}^{-c}$, with $c \sim 0.2$ for $N_{\rm HI} \lesssim 10^{20}$ cm^{-2} and $c \sim 1$ at higher column densities. In our sample, the viral mass and the virial radius grow with time as $\sim a^3$ and $\sim a^2$, and the virial column density decreases as $\sim a^{-1}$. For a self-similar density distribution of neutral hydrogen,

it follows that $c \sim d$, in agreement with the evolution in Figure 2.7.

Comparing next the UVB model (right bars) with the STAR model, a systematic offset towards higher values is apparent from the former. The inclusion of local sources does not drastically alter the covering factor (a factor of two for DLAs and LLSs), but it is clear that ionizing radiation from stars has appreciable effects especially at intermediate column densities and it is required for a consistent comparison with observations (Schaye 2006). The redshift evolution is still present in the UVB model, reassuring us that it is not driven by local sources. Furthermore, this trend cannot be artificially induced by our choice of a constant UVB, since the intensity of the UVB decreases towards the highest redshifts here considered.

In Figure 2.7, we finally compare the covering factor from the streams and galaxies with the streams alone. Overall, the covering factor of highly optically thick gas in the streams is decreasing faster with time than that of galaxies. For the MFP gas, this effect is absent, with streams responsible for the totality of the covering factor at all times. Streams contribute to more than 80% of the LLSs covering factor over the full redshift interval. Instead, a rapid evolution is seen for SLLSs and DLAs. Streams are responsible for as much as 70% of the covering factor at the highest redshifts, but less then 30% of the covering factor at later times as neutral gas accumulates in proximity of the central galaxies an satellites.

Being a normalized quantity, the covering factor within $2R_{\rm vir}$ is suitable for comparisons among different galaxies and redshifts. However, it does not provide direct information on the physical cross section and its evolution. Further, observations probe either the gas cross section or the covering factor within a given projected distance from a source. In the latter case, the most natural quantity to compute is the covering factor within a fixed angular distance from the halo center. In Figure 2.8, we show the mean covering factor within $R_{\rm vir}$ (right columns), $2R_{\rm vir}$ (left columns) and within a circular aperture of 10 arcsec in radius (central column). Again, the number at the top indicates the number of galaxies included in each bin with the second

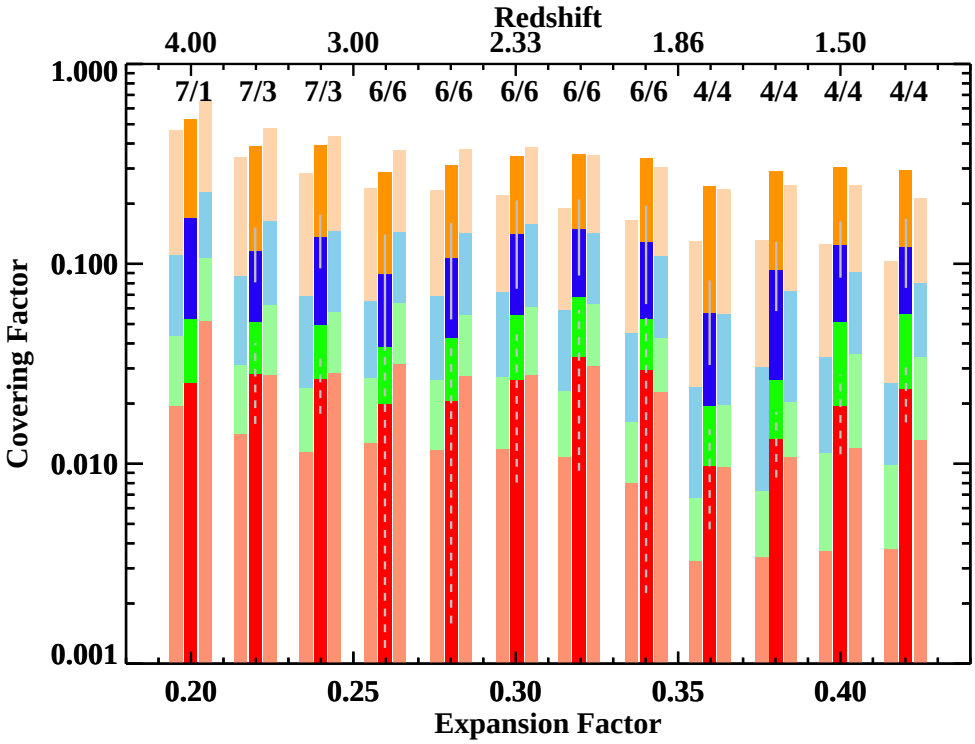

Figure 2.8: Same as Figure 2.7, comparing covering factors that are measured in different areas. In the center, the covering factor is measured within a fixed aperture of 10 arcsec in radius. Shown for comparisons are the covering factor defined within $R_{\rm vir}$ and $2R_{\rm vir}$, on the right and left, respectively. The number of galaxies included in each bin is shown above the columns (variable/fixed apertures) and the standard deviation around the mean is highlighted with solid and dashed lines. The variation with time is much weaker when the aperture is fixed, reflecting the decrease of covering area with distance from the halo center.

number for the fixed aperture. The variance about the mean for LLSs and DLAs is shown with a solid and dashed line, respectively. At the redshifts examined here, an angular distance of 10 arcsec corresponds to $70-85$ proper kpc, equivalent to $\sim 3-5$ times the virial radius at the highest redshifts and to $\sim 0.5-1$ times the virial radius at the lowest redshifts. The snapshots in which the extraction box exceeds the simulated region are not included. Further, for a consistent comparison, we consider a path length through the grid of twice the virial radius in all three cases.

The covering factor measured within $R_{\rm vir}$ is higher than the one inside $2R_{\rm vir}$, particularly for DLAs. This is trivially expected for a gas distribution with declining column density. Considering now a fixed angular distance, the covering factor exhibits a much weaker evolution with a modest increase in the interval $2 < z < 3$, where the sample is more homogeneous. As the universe expands, the mean column density within $2R_{\rm vir}$ drops and the fraction of the area within $2R_{\rm vir}$ subtended by gas above a fixed column density threshold decreases accordingly. However, the virial radius is increasing with respect to the size of the extraction aperture, resulting in a nearly constant or slightly increasing covering factor.

Within $\sim 100/h$ kpc, the covering factor of absorbers in the transverse direction of QSO host galaxies is significantly higher than our estimates, being unity for $N_{\rm HI} > 1.6 \times 10^{17}$ cm^{-2} (Hennawi et al. 2006) and 30% for $N_{\rm HI} > 10^{19}$ cm^{-2} (Hennawi & Prochaska 2007). This discrepancy may originate in the environment around QSOs, more massive than the galaxies in our sample. Further, these simulations do not resolve overdensities on very small scales and consequently the cross section distribution suffers from an additional uncertainty. For example, a population of neutral clouds of $\lesssim 100$ pc in size that are pressure confined by the hot halo gas (see e.g. Dekel & Birnboim 2008, Prochaska et al. 2009a, Birnboim & Dekel 2010) may be missed in these simulations and the cross section of optically thick gas would be underestimated. Similarly, shells or bubbles associated with radiative shocks on small scales may not be properly resolved (cfr. Simcoe et al. 2006).

The weak redshift evolution seen in Figure 2.7 and Figure 2.8 may be the result of having combined galaxies at different mass. To explicitly test this hypothesis, in Figure 2.9 we display the LLS covering factor within $2R_{\rm vir}$ (bottom panel) and within 10 arcsec (top panel) for individual galaxies. Although with significant scatter from object to object, the covering factor within $2R_{\rm vir}$ appears to be roughly constant or slowly declining regardless to the galaxy mass. Similarly, the covering factor within a 10 arcsec aperture is slowly increasing at all masses.

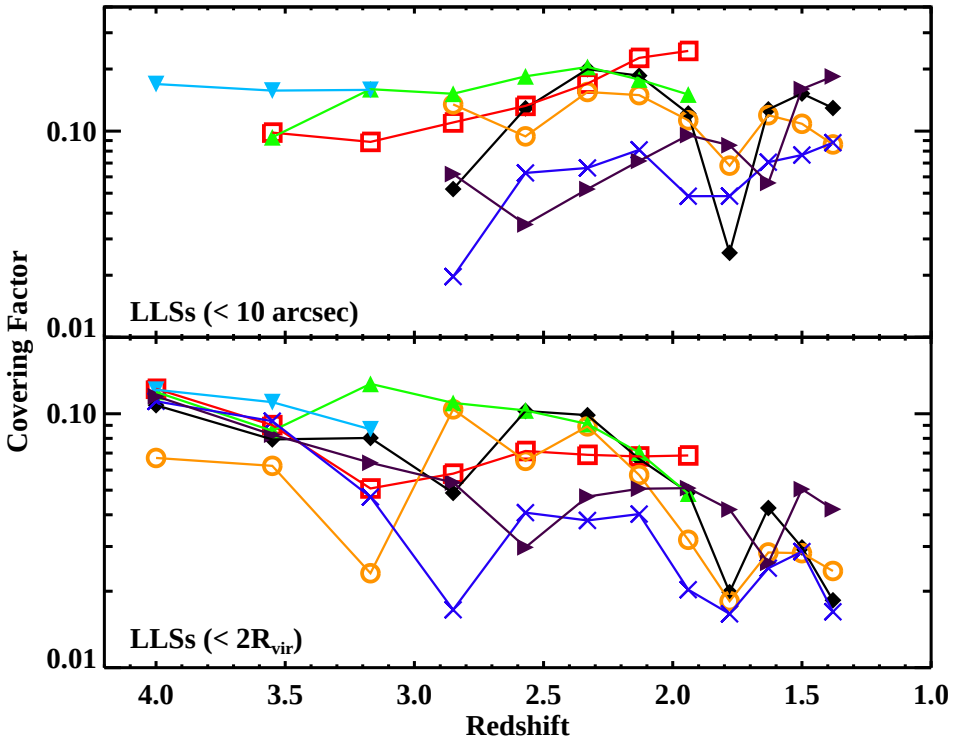

Figure 2.9: Redshift evolution of the covering factor for LLSs in individual galaxies (color coded as in Figure 2.1) within $2R_{\rm vir}$ (bottom panel) and for a fixed aperture of 10 arcsec (top panel). Trends similar to those inferred from Figure 2.7 and 2.8 are visible when inspecting individual galaxies, regardless to their mass.

2.6 Predicted absorption line statistics

In the previous section we discussed the redshift evolution of the covering factor in streams and galaxies, where we found that only a small fraction of the virial area is covered by optically thick gas. In this section, we wish to translate the cross sections and covering factors previously measured into more direct observables that can be used in comparison to real data to uncover signatures of cold gas accretion. We start by introducing some of the formalism adopted in typical measurements of ALSs. Readers familiar with this notation can skip the next sub-section and continue from Section 2.6.2, where we derive the column density distribution function for cold-stream fed galaxies and discuss their visibility in absorption.

2.6.1 Common formalism

To probe any form of gas that lies along the line of sight, observers collect large spectroscopic samples of QSOs that serve as randomly distributed background sources (e.g. Prochaska et al. 2005, O'Meara et al. 2007). In this experiment, there is no *a priori* knowledge of the foreground system (see however Rafelski et al. 2011) and one can directly measure the probability to intersect gas along a random sightline. This quantity translates to a sky covering factor that is the percentage of total observed area covered by gas.

One can define the incidence

$$\ell(z) = \frac{N_{\rm abs}}{\Delta z} \tag{2.3}$$

as the number of systems $N_{\rm abs}$ detected across the total searched redshift Δz. This quantity is related to the probability of intersecting a system along the line of sight at redshift z. In analogy with the idealized experiment of a body moving in a sea of particles with number density n and

cross section σ, the following equality holds

$$\ell(z)dz = n_{\rm abs}\sigma_{\rm abs}dl_{\rm c} \tag{2.4}$$

with $n_{\rm abs}$ the comoving number density of absorbers, $\sigma_{\rm abs}$ their comoving cross section and $dl_{\rm c}$ the infinitesimal comoving path length.

Within a given cosmology $H(z) = H_0[\Omega_\Lambda + (1+z)^3\Omega_m]^{1/2}$ and $dl_{\rm c} = c/H(z)\ dz$. Replacing z by an "absorption length" X

$$dX = \frac{H_0}{H(z)}(1+z)^2 dz \tag{2.5}$$

and using the identity $\ell(z)dz = \ell(X)dX$, the incidence can be rewritten as the number of systems per absorption length (Bahcall & Peebles 1969)

$$\ell(X) = \frac{c}{H_0} n_{\rm abs}\phi_{\rm abs} \tag{2.6}$$

where the cross section $\phi_{\rm abs}$ now represents the physical area of the absorber.

Equation (2.6) is the zeroth moment of the column density distribution function $f(N_{\rm H\,I}, X)$ that characterizes the number of ALSs per absorption length as a function of the H I column density. From the definition of $\ell(X)$, it follows that

$$f(N_{\rm HI}, X) = \frac{c}{H_0} n_{\rm abs}(N_{\rm HI}, X) A_{\rm abs}(N_{\rm HI}, X)\ , \tag{2.7}$$

with $A_{\rm abs}$ the physical area per unit $N_{\rm HI}$. Equation (2.7) reveals that absorption line statistics directly probe the shape and the evolution of the physical cross section times the number density of absorbers rather than the covering factor of gas around these systems.

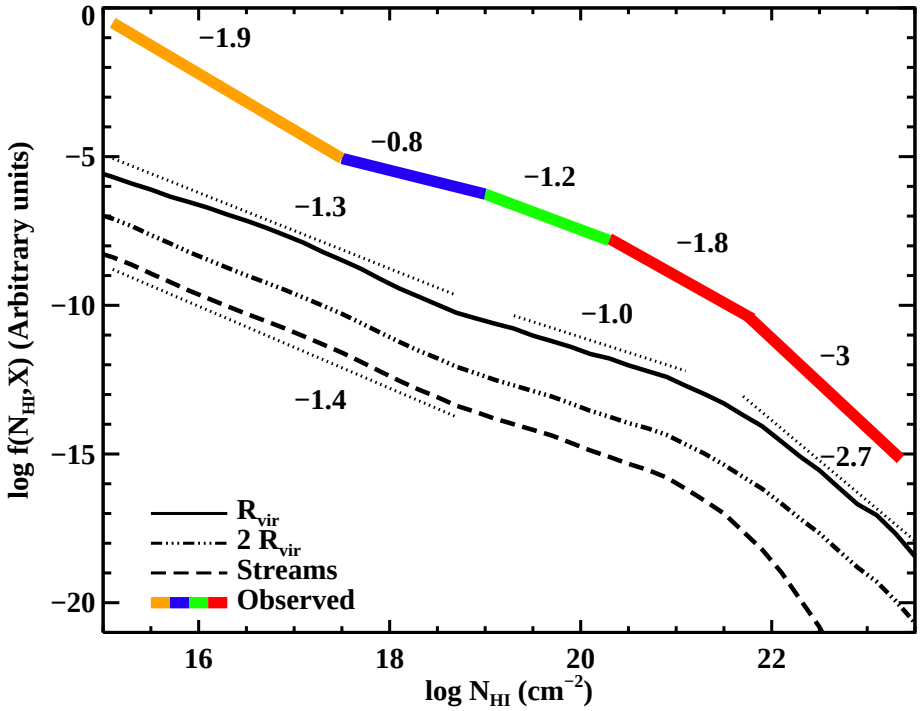

Figure 2.10: Comparison of the shape of the simulated and observed column density function $f(N_{\rm H\,I}, X)$. The normalizations of the different curve are arbitrary, to highlight the shapes only. The simulated $f(N_{\rm H\,I}, X)$ from the STAR model at $z \sim 3.5$ is shown for gas inside the virial radius (solid line), for gas within $2R_{\rm vir}$(dash-dotted line), and with for the streams only (dashed line). The observed $f(N_{\rm H\,I}, X)$ at $z \sim 3.7$ from Prochaska et al. (2010) is shown as a colored curve. Power law regressions to the simulation results in intervals of column density are shown as dotted lines. The numbers denote the power law indexes. Simulations of massive galaxies at high resolution reproduce the observed cutoff at high column density, but predict an $f(N_{\rm H\,I}, X)$ that is shallower than observed around $N_{\rm HI} \sim 10^{20}$ cm^{-2} and for SLLSs. Cold streams dominate the $f(N_{\rm H\,I}, X)$ below $N_{\rm HI} \sim 10^{18}$ cm^{-2}, but a larger contribution from the intergalactic medium is required at lower $N_{\rm HI}$ to satisfy the constraints imposed by the observed mean free path of ionizing radiation.

2.6.2 The column density distribution function

To understand the nature of the gas probed in absorption, it is common practice to reproduce the $f(N_{\rm H\,I}, X)$ in its shape and normalization from simulations of large volumes, required for statistical significance. Here we take a different approach where we derive an $f(N_{\rm H\,I}, X)$ that, by construction, is due to gas associated only with massive galaxies and the streams with merging galaxies that feed them. Any similarity or discrepancy with the observed distribution function informs the features of $f(N_{\rm H\,I}, X)$ that reflect the gas distribution in these particular class of systems as predicted by these simulations. The reader should keep in mind that our sample and our $f(N_{\rm H\,I}, X)$ are not intended to be representative of the full ALS population.

Our model for $f(N_{\rm H\,I}, X)$ follows the definition in Equation (2.7). The physical area per unit $N_{\rm HI}$ is constructed from the average cross section distribution function over all the galaxies at a common redshift

$$\overline{A_{\rm gal}(N_{\rm HI}, z)} = \sum_{i}^{N_{\rm gal}} \frac{\phi_{\rm gal,i}(N_{\rm HI}, z)}{N_{\rm gal} \Delta N_{\rm HI}} \, . \tag{2.8}$$

We consider the virial radius to define the gas that is physically associated with a galaxy, but we construct also a second model considering the cross section within $2R_{\rm vir}$ to stress the effects of the cold streams. For the comoving number density of absorbers, we use the cumulative number density[4] of galaxies that are more massive than the least massive galaxy at a given redshift. In practice, we construct an average cross section that we attribute to all the halos of equal mass or more massive than the galaxies in our sample. This procedure implies an extrapolation towards higher masses, since systems much above 10^{12} $\rm M_{\odot}$ are not represented in our sample. Due to

[4]The comoving number density is computed following the approximations described in Dekel & Birnboim (2006; Appendix A). Recent simulations (e.g. Klypin et al. 2011) reveal deviations from these approximations, but they become relevant only above $z \sim 4$.

the scatter from galaxy to galaxy in the cross section and to its mass dependence, $\overline{A_{\rm gal}(N_{\rm HI}, z)}$ and the final $f(N_{\rm H\,I}, X)$ should be regarded as approximations.

In Figure 2.10 we compare the functional form of the $f(N_{\rm H\,I}, X)$ derived for the STAR model at $z \sim 3.5$ within the virial radius (solid black line), within twice the virial radius (dashed-dotted line), and in the streams alone (dashed line). The observed distribution at $z \sim 3.7$ from Prochaska et al. (2010) is superimposed with colors. For DLAs and SLLSs (i.e. $N_{\rm HI} > 10^{19}\,{\rm cm}^{-2}$ where one resolves the damping wings of Lyα), $f(N_{\rm H\,I}, X)$ is derived from observations of individual absorbers, but at lower column densities observations provide only integrated constraints on the distribution function. For this reason, the shape of $f(N_{\rm H\,I}, X)$ in the LLSs and MFP region is more uncertain. In Figure 2.10, we report the preferred model from Prochaska et al. (2010).

Although not a perfect match, these models reproduce some of the features of the observed $f(N_{\rm H\,I}, X)$. Three different intervals can be identified. First, there is a steep decrease at high column density with an evident break around $N_{\rm HI} = 3 \times 10^{21}$ cm^{-2}. From our model, we measure a power law ~ -2.7 for $\log N_{\rm HI} \gtrsim 21.7$ (in cm^{-2}; dotted line), similar to the observed slope ~ -3. This cutoff naturally emerges from our high resolution simulations of massive disks, while it is hard to reproduce in cosmological simulations at lower resolution (e.g. Pontzen et al. 2008, Tescari et al. 2009, Cen 2012; but see Altay et al. 2010, McQuinn et al. 2011). The inclusion of molecules would suppress even further the number of systems with H I above 10^{22} cm^{-2}(see Figure 2.17), bringing the model to even better agreement with observations.

Moving towards lower column densities, the model exhibits a plateau with slope ~ -1 between $N_{\rm HI} = 3 \times 10^{18} - 10^{21}$ cm^{-2}. The observed $f(N_{\rm H\,I}, X)$ distribution also flattens below 10^{20}cm^{-2} (O'Meara et al. 2007, Prochaska et al. 2010) which has been interpreted as the result of transitioning from predominantly neutral gas to predominantly ionized (Zheng & Miralda-Escudé 2002, McQuinn et al. 2011). In fact, our model flattens at somewhat high $N_{\rm HI}$ and

underpredicts the observed incidence of SLLSs relative to DLAs. Provided that SLLSs arises from galaxies, perhaps this discrepancy would be resolved by including less massive galaxies than those modeled here. As previously noted for the cross section, we might be missing a population of small clouds of neutral hydrogen due to limited resolution. Finally, below $N_{\rm HI} = 10^{18}$ cm^{-2} the distribution function becomes moderately steeper with slope ~ -1.3. This qualitatively follows the observed distribution, but it is not sufficiently steep to reproduce the integrated mean free path to ionizing radiation at $z \sim 4$ (Prochaska et al. 2009b). We infer, therefore, that gas beyond the virial radius must contribute significantly at these column densities (see also Kohler & Gnedin 2007). As already seen in Figure 2.4, due to the stream contribution (dashed line) the $f(N_{\rm H\,I}, X)$ steepens mildly below $N_{\rm HI} \sim 10^{18}$ cm^{-2}, but gas within $2R_{\rm vir}$ is not enough to reproduce the observed slope (dash dotted line). Contribution from the IGM at larger radii may be required.

Figure 2.11 compares again the observed $f(N_{\rm H\,I}, X)$ with the model inside the virial radius, including at this time the absolute normalization. The observed $f(N_{\rm H\,I}, X)$ has been corrected to match our adopted cosmological parameters. Values are listed in Table 2.5. This figure reveals an overall satisfactory agreement between the observations and the models for high column density DLAs and at $N_{\rm HI} \sim 10^{18}$ cm^{-2}. As already noted, the two distributions diverge entering the optically thin regime, suggesting a missing IGM contribution in the simulations. Also missing are SLLSs and high column density LLSs that are not found in large enough numbers within the halos of massive galaxies. The UVB model, superimposed as a dashed line, has a larger fraction of LLSs and SLLSs, but not enough to account for the missing systems. It is interesting to note that in these models SLLSs originate from gas at the outskirts of the disks and are ionized by the leaking UV photons from stars that are forming in the disk (Schaye 2006). While most of the $f(N_{\rm H\,I}, X)$ is shaped by the UVB (Razoumov et al. 2006, Nagamine et al. 2010), RT effects from local sources produce non-negligible effects. Albeit with all the

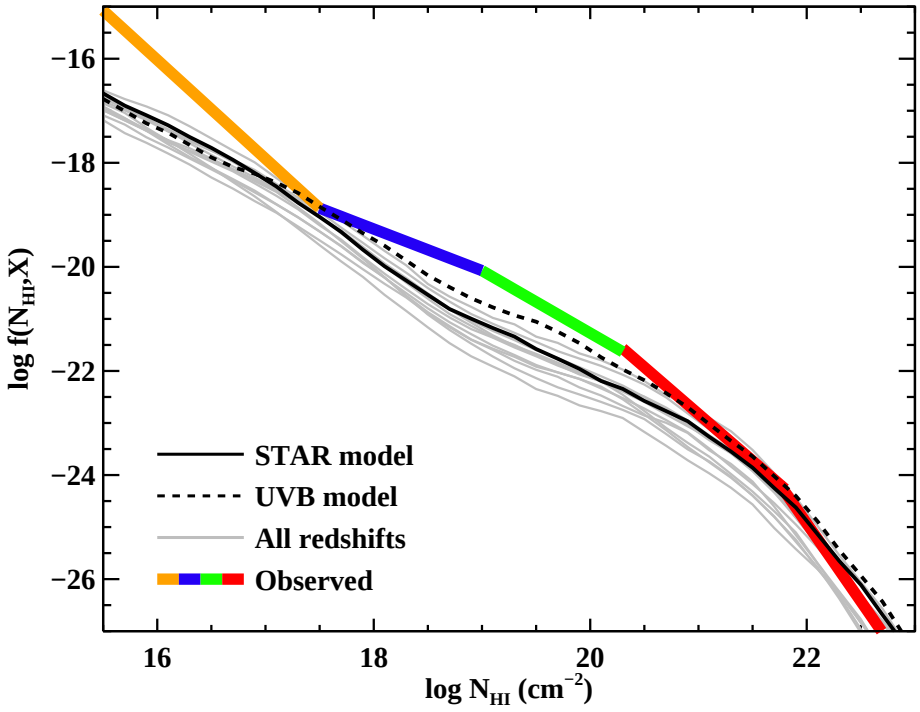

Figure 2.11: Column density distribution function from observations at $z \sim 3.7$ (colors; Prochaska et al. 2010) and simulations for the STAR (solid line) and the UVB (dashed line) models at $z = 3.5$ within the virial radius. Shown in gray are simulated $f(N_{\rm H\,I}, X)$ for different redshifts in the interval $z = 4 - 1.4$. There is a satisfactory agreement between theory and observations for DLAs and at $N_{\rm HI} \sim 10^{18}$ cm^{-2}, while the two distributions diverge at lower column densities. SLLSs and high density LLSs are underrepresented by these models. The $f(N_{\rm H\,I}, X)$ exhibits modest redshift evolution in shape, in agreement with observations.

uncertainties discussed, from this simple comparison we infer that massive disks shape the $f(N_{\rm H\,I}, X)$ at the highest column densities and cold streams contribute non negligibly to the LLS population, particularly between $N_{\rm HI} \sim 10^{17} - 10^{18}$ cm^{-2}.

Finally, in gray, we superimpose different $f(N_{\rm H\,I}, X)$ for the redshift interval $z = 4 - 1.4$. These models are remarkably invariant with redshift and this limited evolution in the $f(N_{\rm H\,I}, X)$ is consistent with the observation that the column density distribution function of DLAs preserves its shape over 12 Gyr (Prochaska & Wolfe 2009). This is also consistent with the present day distribution function from 21-cm observations in H I rich disks. Approaching the optically thin regime, a larger degree of evolution is observed (Prochaska et al. 2010, Ribaudo et al. 2011), which is attributed to radiative transfer effects. This variation is not captured by our model, probably because massive galaxies are less sensitive to the effect of the UVB than the IGM and low mass galaxies which are needed in order to fully match the observed distribution.

The limited evolution in the models presented in Figure 2.11 can be qualitatively understood from visual inspection of Figure 2.6. As the virial radius grows with time, the gas preserves its distribution with centrally concentrated neutral regions surrounded by SLLSs. In turn, this column density interval marks the transition to more filamentary structures that dominate the LLSs and MFP cross section. More quantitatively, this non-evolution is related to the fact that the neutral hydrogen column density PDF is found to be nearly self-similar. Therefore, in our simulations, the lack of evolution in $f(N_{\rm H\,I}, X)$ is driven by the distribution of gas within each halo (i.e. the cross section) more than by the cosmological evolution (i.e. the number density of systems).

2.6.3 Are cold streams visible in absorption?

By integrating the column density distribution function, one can compare predictions from the simulations with the observed incidence of absorbers. Figure 2.11 shows that the nor-

malization of the simulated $f(N_{\rm H\,I}, X)$ is not too far from the observed one. In these simulations, massive galaxies ($M_{\rm vir} \sim 10^{10} - 10^{12}$ M$_\odot$) alone can account for $\sim 20\% - 30\%$ of the observed DLA and LLS population. The inclusion of satellites and streams up to twice the virial radius further increases this contribution (cfr. Maller et al. 2003). Although uncertain up to a factor two, these numbers suggest that a non-negligible fraction of observed ALSs originate in galaxies and streams as represented in these simulations.

This result is not at odds with the low covering factor for optically thick gas discussed in the previous sections. This is because the incidence of absorbers with background QSOs is sensitive to the physical cross section times the number density of halos. Given a large enough number density of halos, even galaxies with relatively small covering factor can offer enough total cross section to account for a significant fraction of the observed incidence. Most of the DLAs are estimated by theory to originate in the mass interval $10^9 - 10^{11}$ M$_\odot$ (e.g. Pontzen et al. 2008, Tescari et al. 2009), with a peak around $\sim 5 \times 10^{10}$ M$_\odot$. In models with strong feedback (e.g. Nagamine et al. 2007, Cen 2012), the peak shifts to even higher masses. It is not surprising that our sample that includes galaxies with $5 \times 10^{10} - 3 \times 10^{11}$ M$_\odot$ at $z \sim 3$ is within a factor of a few of the observed incidence.

Combining the observed incidence for DLAs and LLSs with the cross section from our simulations, we can infer the minimum halo mass required to fully match observations. At $z \sim 3.5$ ($z \sim 2.8$), the average cross section within $2R_{\rm vir}$ is ~ 304 kpc^2 (~ 530 kpc^2) for DLAs and ~ 1548 kpc^2 (~ 2010 kpc^2) for LLSs. The halo number density required to match the observed incidence of DLAs and LLSs (Prochaska & Wolfe 2009, Prochaska et al. 2010, Ribaudo et al. 2011) is ~ 0.062 Mpc^{-3} (~ 0.032 Mpc^{-3}), corresponding to a minimum halo mass of 1.6×10^{10} M$_\odot$ (3.6×10^{10} M$_\odot$). This is a factor 1.4 (2.7) below the minimum mass included in this sample. This calculation is very crude since, for example, at lower masses the mean cross section is expected to decrease (e.g. Figure 2.5). Further, these results are sensitive to

the radius adopted for the cross section determination, here chosen to be $2R_{\rm vir}$, i.e. the size of the re-simulated box. Since ideally one would adopt the mean separation between halos, larger than $2R_{\rm vir}$ at these masses, our numbers may be even underestimated. With these limitations, our calculation suggests that only a small extrapolation towards lower masses is needed to fully account for the observed incidence of ALSs.

In summary, despite their small covering factor in each halo, the cold streams should have already been detected in absorption in large spectroscopic surveys. Since most of the cross section in cold streams is predicted to be at $N_{\rm HI} \lesssim 10^{19}$ cm^{-2}(see Figure 2.4 and Figure 2.10), LLSs are the best candidates for the cold gas in the inflowing streams. If there are other components of cold gas in the halo that may not exist in proper abundances in our current simulations (e.g. high velocity cloud analogs or massive outflows), then other indicators are required to distinguish between them and the cold streams.

2.7 Observable properties of the CGM

In a second class of experiments designed to probe cold gas in absorption, observers select an object or a class of objects of interest and use background sources to probe the gas distribution around the foreground systems. Using a background quasar (or galaxy) and knowing the angular separation between the probe and the object of interest, this experiment yields a direct measurement of the angular H I distribution (e.g. Christensen et al. 2007, Monier et al. 2009, Fumagalli et al. 2010b) about the object of interest or of the gas covering factor and kinematics (e.g. Steidel et al. 2010). In the next section, we provide a summary of the formalism adopted in the analysis of absorption lines as well as a more technical description on how line profiles are derived from these simulations. Starting from Section 2.7.2, we discuss predictions for the metallicity and kinematics of absorption lines, providing a comparison with

recent observations.

2.7.1 Common formalism and numerical procedures

2.7.1.1 The equivalent width

The normalized profile of an absorption line is characterized by the equivalent width

$$W(\lambda_0) = \frac{\lambda_0}{c} \int_{-\infty}^{+\infty} \left[1 - \bar{I}(v)\right] dv \,, \tag{2.9}$$

with $\bar{I}$ the normalized intensity and λ_0 the rest frame wavelength at the line center. The equivalent width provides a direct measurement of the light absorbed along the line of sight. Through the radiative transfer equation, $W(\lambda_0)$ can be related to the gas optical depth that in turn depends on the gas column density and kinematics and, via the oscillator strength, on the quantum mechanical properties of the transition.

Three different regimes are typically identified in the so-called curve of growth, that describes the variation of the equivalent width with both column density and oscillator strength. At very high column densities, in the damped regime, all the light at the line center is absorbed and the equivalent width is a solely function of the column density. At intermediate column densities, on the saturated portion of the curve of growth, the line center is again fully black, but the equivalent width is most sensitive to the line of sight component of the gas velocity that regulates the line width. Finally, at low column densities, in the linear regime, the equivalent width becomes proportional to the column density in the observed direction and the absorption profile is no longer zero at the line-center. Averaging along different sightlines, one can further probe the mean covering factor around a class of sources by looking at the fraction of transmitted light close to the line center.

2.7.1.2 Techniques for the simulated line profiles

To reproduce the absorption line-profile for Lyα and Si IIλ1260, we compute the transmitted intensity I_ν of a background source with intensity I_0 in a random projected sightline i

$$I_{\nu,i} = I_0 \exp(-\tau_{\nu,i}) \,, \tag{2.10}$$

where the optical depth $\tau_{\nu,i}$ is given by

$$\tau_{\nu,i} = \sum_{k|i} N_k s \phi_k(\nu) \,. \tag{2.11}$$

Here, the sum is over all the AMR cells aligned along the i-th line of sight, and N_k is the column density in each cell. The frequency integrated absorption cross section s is given by

$$s = \frac{\pi e^2}{m_e c} f \,, \tag{2.12}$$

with f the oscillator strength, e the electron charge, m_e the electron mass and c the speed of light. The frequency-dependent line-profile is

$$\phi_k(\nu) = \frac{H_k(u,a)}{\Delta\nu_{D,k}\sqrt{\pi}} \,, \tag{2.13}$$

with the Voigt function

$$H(u,a) = \frac{a}{\pi} \int_{-\infty}^{+\infty} dy \frac{\exp(-y^2)}{(u-y)^2 + a^2} \,. \tag{2.14}$$

Here, $a = \gamma/(4\pi\Delta\nu_D)$, $u = (\nu - \nu_l)/\Delta\nu_D$, $y = v/b$ and $\Delta\nu_D = (\nu_0 b)/c$. In the above equations, $b^2 = (2kT/m_X + \zeta^2)$ is the broadening parameter given by that gas temperature and element mass, γ is the sum over the spontaneous emission coefficients, ν is the observed spectral fre-

quency, and $\zeta = 10$ km/s the turbulent velocity. We assume the central galaxy is at zero velocity (corresponding to the rest-frame frequency ν_0) and we compute for each gas cell the line center frequency ν_l corrected for the gas velocity along the line of sight. We integrate Equation (2.14) using the formalism provided in Zaghloul (2007).

All the spectra are computed at a resolution of 4 km/s. For Lyα, we assume $f = 0.4164$ and $\gamma = 6.265 \times 10^8$. For Si II$\lambda$1260, we use $f = 1.007$, and $\gamma = 2.533 \times 10^9$ and we derive the Si II density from the neutral hydrogen density and metallicity, assuming $\log(\mathrm{Si/H})_\odot + 12 = 7.51$ (Asplund et al. 2009). To compensate for the different ionization potential of silicon and hydrogen, we boost the Si II volume density by a factor of 10 in cells with $x_{\mathrm{HI}} < 0.1$. The amplitude of this boost depends on the gas and radiation field properties, but it will become clear from our analysis that a variation of several order of magnitudes in the Si II column density is necessary to significantly alter our results. In this simple model, we also neglect Si II recombination at low temperature and a lower collisional rate in shielded regions. These effects alter the Si II column density, but only in regions of small covering factor with minor consequences on the average absorption lines. A complete analysis of metal lines requires the inclusion of collisional ionization and photoionization, which is beyond the scope of this paper.

2.7.2 Metallicity

The metal enrichment of the gas may help distinguish between gas that is accreting from the cosmic web and the more enriched material that is located in proximity to a galaxy or is outflowing from regions with intense star formation. Production and diffusion of metals is known to be problematic in numerical simulations (Wadsley et al. 2008), but the AMR code used for our simulations should minimize problems of metal transport typical of SPH simulations (but see Shen et al. 2010). Caution is advised in generalizing these comparisons of metallicity between our simulations and observations because our current simulations are limited to supernova-

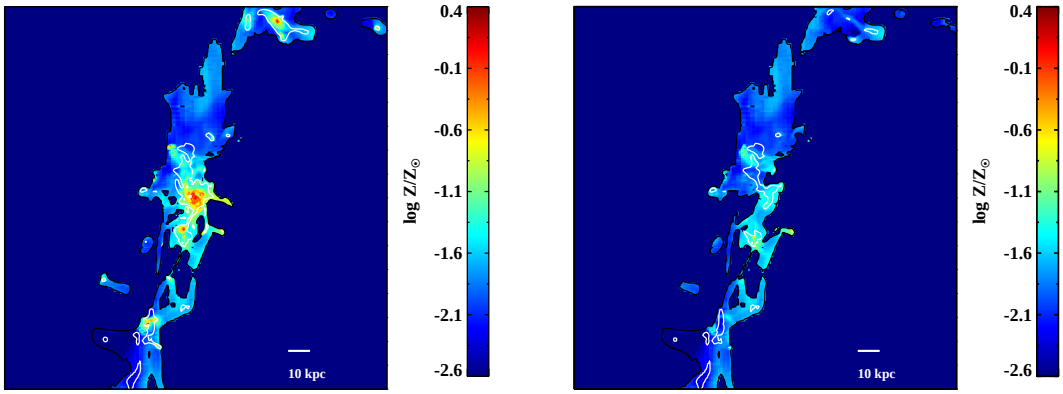

Figure 2.12: Metallicity in the optically thick gas of MW3 at $z = 2.3$, weighted by hydrogen column density. Top: galaxies and streams. Bottom: streams alone. The black and white contours denote $\log(N_{HI}/cm^{-2}) = 17.2$ and $\log(N_{HI}/cm^{-2}) = 20.3$ respectively. A metallicity gradient is visible. Gas in and around galaxies is more enriched than in the streams alone.

driven winds which do not reproduce outflows as massive as suggested by some observations (e.g. Steidel et al. 2010), and also because the specific IMF Miller & Scalo (1979) is assumed in the simulations.

In Figure 2.12, the column density weighted metallicity in solar units ($Z_\odot = 0.0134$; Asplund et al. 2009) for a projection of MW3 at $z \sim 2.3$ is shown. On the top, we display metals in both streams and galaxies for optically thick gas; on the bottom, we present the metallicity in the streams alone. Contours are for $N_{HI} = 1.6 \times 10^{17}$ cm^{-2}(black line) and $N_{HI} = 2 \times 10^{20}$ cm^{-2}(white line). A spatial gradient in metallicity is visible. DLA gas inside the main galaxy and satellites is significantly enriched, with values close to 1/10 solar (in reasonable agreement with DLA observations; e.g. Prochaska et al. 2003). The fraction of metals decreases moving far from the central disks and at lower column densities. The bulk of the LLSs is enriched at $1/10 - 1/100$ solar, with the streams skewed at the lowest metallicity.

A similar metal content characterizes the entire sample at $z = 1.3 - 4$, almost independently of redshift and mass. Figure 2.13 shows the distribution of column density weighted metallicity ψ in DLAs (red dashed line), SLLSs (green dotted lines) and LLSs (blue dash dotted

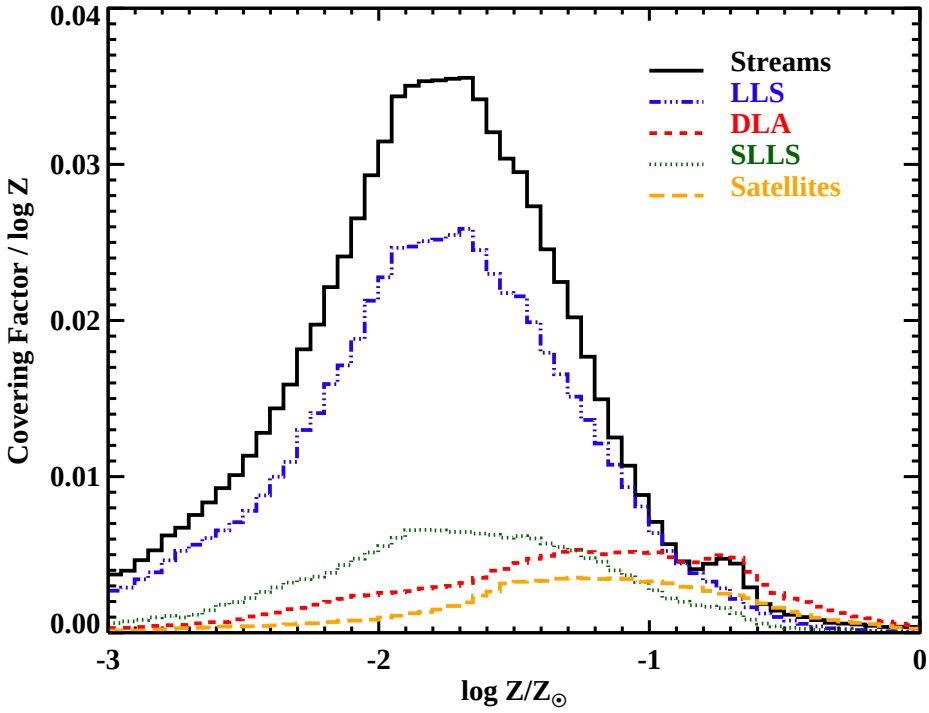

Figure 2.13: Distribution of metallicity, column density weighted, in the entire sample for DLAs (red dashed line), SLLSs (green dotted line) and LLSs (blue dot-dashed line). Also shown are the streams alone (black solid line) and satellites without centrals (orange long dashed line). The smooth stream component is more metal poor than the galaxies, with typical $Z \sim 0.01 Z_{\odot}$. However, gas in streams is not pristine, but already enriched by previous episodes of star formation.

line). The metal distributions for streams alone (only for $N_{\rm HI} > 1.6 \times 10^{17}$ cm^{-2}; solid line) and for satellites without centrals (orange long-dashed line) are superimposed. This distribution has been normalized so that the integral over the metallicity gives the average covering factor in the entire sample for a fixed range of column density

$$\psi(\Delta N_{\rm HI}, Z) = \frac{\phi(\Delta N_{\rm HI}, \Delta Z)}{\Delta z \ \phi_{\rm tot}} \, . \tag{2.15}$$

Here, ϕ is the physical cross section.

From Figure 2.13, we see that most of the area that is covered by optically thick gas is occupied by metal poor cold filaments, with a distribution centered at $Z \sim 10^{-2} Z_{\odot}$ and a spread of over ~ 1 dex. Cold streams are responsible for a metal poor population of LLSs, while optically thick absorbers that originates within $0.25 R_{\rm vir}$ of centrals and satellites can be enriched above $\log Z/Z_{\odot} = -1$. DLAs are characterized by a broad distribution of metals centered at $\log Z/Z_{\odot} \sim -1$ with a tail to lower metallicity. Satellites alone have a distribution similar to the one of DLAs. SLLSs originate from ionized material in the surroundings of the disks and have an equally broad distribution, centered around $\log Z/Z_{\odot} \sim -1.5$. Outflows more vigorous than the ones generated in our simulations are expected to produce even higher metallicity around the star-forming regions. Since more primordial gas that replenishes galaxy disks with fresh fuel for star formation is predicted to have one order of magnitude less metals than gas in galaxies, we conclude that metal poor LLSs in proximity to galaxies are the best candidates for cold gas in inflowing streams. The metallicity distribution in LLSs (e.g. Prochter et al. 2010) may offer a crucial test to distinguish between cold inflowing streams and gas that already resides in the galaxies or is outflowing (see also Giavalisco et al. 2011).

2.7.3 Absorption line profiles

The information on the metallicity and covering factor can be combined together with kinematics to formulate predictions on the strength and shape of absorption line profiles at different projected distances from the galaxy centers. These models can then be compared with the population of ALSs observed in the CGM of Lyman break galaxies at $z \sim 2-3$ (Steidel et al. 2010). We perform this calculation by stacking several spectra generated along multiple sightlines in a subset of our sample (MW1, MW2, MW3 and MW8 in two projections) at $z \sim 2.3$ and $z \sim 3.2$. The absorption profiles at $z \sim 2.3$ for Lyα (top panel) and Si IIλ1260 (bottom panel) are shown in Figure 2.14, where each spectrum is a composite of independent sightlines

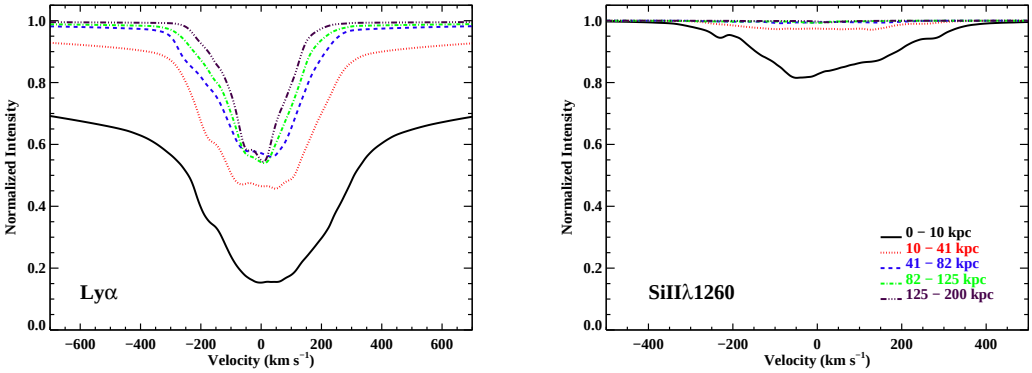

Figure 2.14: Averaged Lyα (top) and Si IIλ1260 (bottom) absorption lines for background sources at different impact parameters from the central galaxy. The Lyα profiles are never saturated due to the low covering factor and are characterized by a FWHM of $\sim 200-300$ km/s for $b > 41$ kpc. The low covering factor, low intrinsic Si/H abundance and low metallicity produce very weak average Si IIλ1260 absorption.

in five intervals of impact parameter, chosen to match the observations presented in Steidel et al. (2010).

Despite the fact that Lyα saturates at low H I column densities, none of the spectra are black at the line-center due to non-unity covering factor of neutral gas. The transmitted normalized intensity at the line center ranges in the interval $0.5-0.6$, similar to what is observed. Damping wings are visible within the innermost 40 kpc, where most of the highly optically thick gas resides. The line profiles are characterized by a full width at half maximum (FWHM) velocity of $\sim 400-600$ km/s for $b < 41$ kpc, related to the kinematics of the central galaxies and the inner satellite relative motions, as well as the bulk velocity of the incoming streams. Conversely, at larger impact parameters ($b > 41$ kpc), a velocity of $200-300$ km/s FWHM is most likely associated with the inflowing gas that is streaming at ~ 200 km/s or more (Dekel et al. 2009b). The line profiles are symmetric, despite the fact that the gas is infalling towards the central galaxies, indicating that there are no preferred directions in the velocity field in a stack of multiple sightlines. Inspecting the metal lines, it is evident that the low covering factor

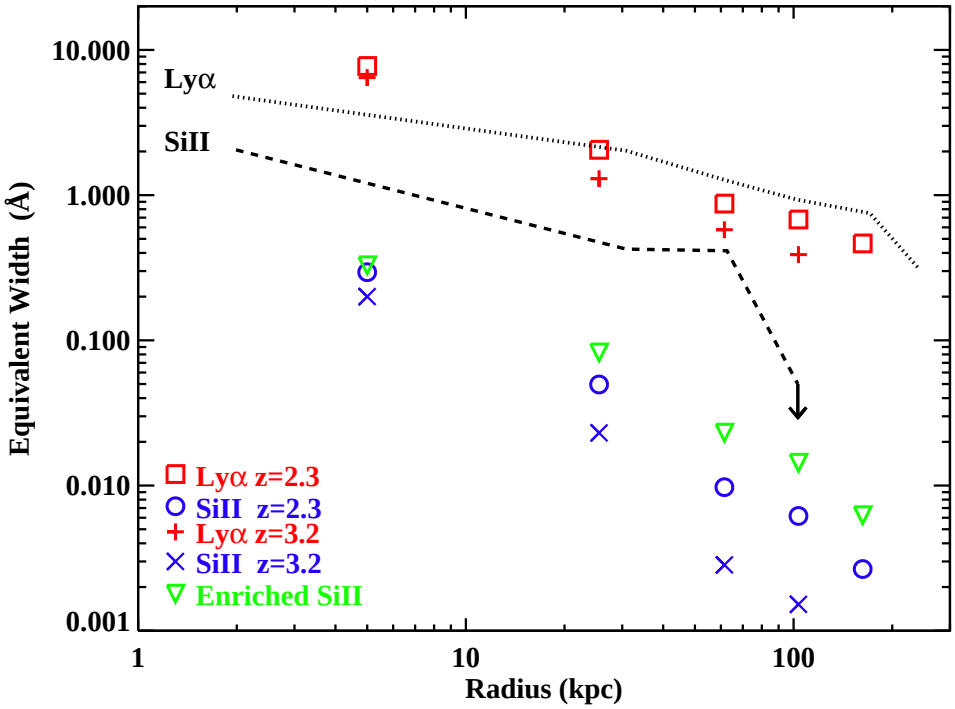

Figure 2.15: Equivalent width profile at $z \sim 2.3$ and $z \sim 3.2$ for Lyα (red open squares and crosses) and Si IIλ1260 (blue open circles and x's). The black dotted and dashed lines show the observed rest frame equivalent widths in LBGs at redshift $z \sim 2-3$ (Steidel et al. 2010). Massive galaxies that are accreting gas through cold streams have a large enough covering factor and line-of-sight velocity to account for the observed strength of the Lyα absorption. Conversely, the metal poor streams cannot reproduce the observed equivalent width in metal lines. Boosting the degree of enrichment by introducing an artificial metallicity floor at $0.1Z_{\odot}$ (green downward triangles) is not enough for matching the observations.

and the intrinsic low Si/H abundance, together with the low metallicity of the streams, produce absorption lines that are weaker than the observed values at all impact parameters $b > 10\,\mathrm{kpc}$.

In Figure 2.15 we present a comparison between the line strength from simulations and observations. For each radial bin, the rest frame equivalent width of Lyα at $z \sim 2.3$ and $z \sim 3.2$ are shown with red open squares and crosses, respectively. Similarly, we superimpose the equivalent width of Si IIλ1260 at $z \sim 2.3$ (blue open circles) and $z \sim 3.2$ (blue x's). Values are listed in Table 2.7. The black dotted and dashed lines are for the observed rest frame

equivalent width from spectra of LBGs at redshift $z \sim 2-3$ (Steidel et al. 2010). As expected from direct inspection of the absorption profiles, massive galaxies that are accreting gas through cold streams have a large enough covering factor and sufficiently complex kinematics in H I gas to match observed Lyα values. We emphasize that this is achieved through the gas associated with satellites and streams alone; the supernova-driven winds in our current simulations are hot and not as strong as the outflows suggested by Steidel et al. (2010). While winds, commonly observed in star-forming galaxies, are essential ingredient for realistic models, the fact that the same observations can be reproduced by two extreme scenarios stands as a reminder that evidences for outflowing/inflowing gas are subtle and more refined models are now required.

That these simulations fail to reproduce the strength of Si II at all impact parameters $b < 100\,\mathrm{kpc}$ is not surprising since, due to the lack of strong outflows of cold gas, low-ion metals are tracing the distribution of hydrogen in the galaxies and streams. We can consider whether, for example, the discrepancy for Si IIλ1260 might be attributable to the low enrichment of the cold streams. In Figure 2.16, we show the ratio of the Si II to Lyα optical depth, integrated in the velocity interval $\pm$250 km/s. Only gas with $N_{\mathrm{HI}} > 10^{13}$ cm^{-2} is displayed. The low abundance of silicon relative to hydrogen at solar metallicity suppresses the metal line opacity by five orders of magnitude and the low metallicity of cold streams decreases the Si II opacity even further, by a factor of $\sim 10-100$. For these reasons, metal lines that originate in cold streams are weak and likely to be undetected. A higher degree of enrichment is not enough to boost the equivalent width by one order of magnitude, as demonstrated by a simple experiment (see also Kimm et al. 2011). If we artificially impose a floor in metallicity at $Z = 0.1Z_{\odot}$ for all the gas with $N_{\mathrm{HI}} > 10^{15}$ cm^{-2}, we obtain the equivalent widths shown by the green downward triangles in Figure 2.15. A much higher degree of enrichment increases the line strength only by a modest value.

We conclude that cold streams are unlikely to produce the large equivalent widths of

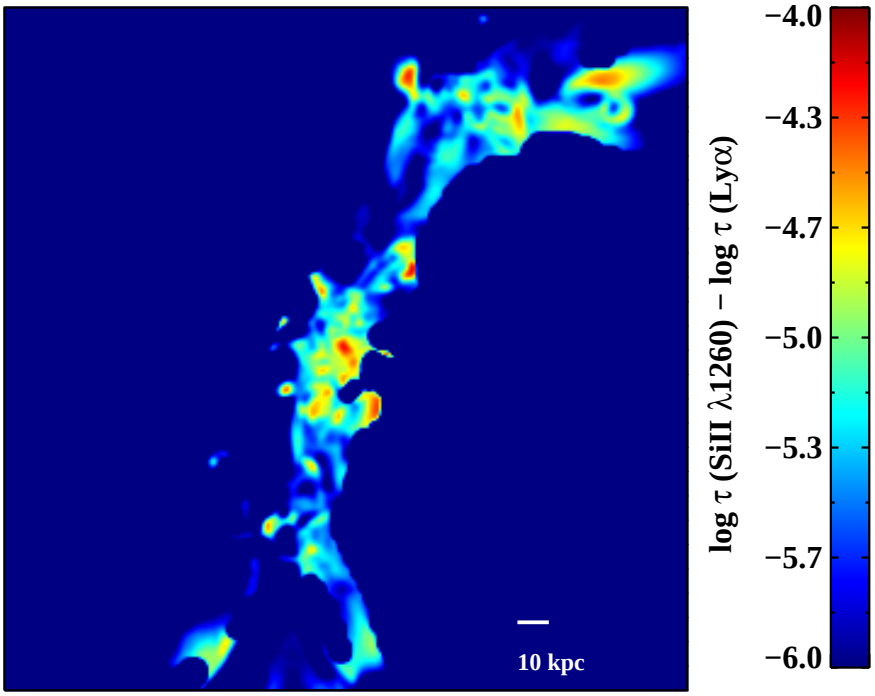

Figure 2.16: Optical depth integrated over velocities in the interval ±250 km/s for Si IIλ1260 relative to Lyα for gas with $N_{\rm HI} > 10^{13}$ cm^{-2}. The low intrinsic Si/H abundance and metal content of cold streams suppress the metal line opacity by more than five orders of magnitude compared to Lyα.

low-ion metal absorption around massive galaxies. If the average equivalent widths reported by Steidel et al. (2010) are confirmed (e.g. with higher resolution observations), an alternate source of opacity must be included in the simulations. This could include, for example, absorption by neighboring galaxies (not modeled here), small clumps of cold gas embedded in a hotter medium, and/or galactic-scale outflows. However, in order to satisfy the constraints imposed by the Lyα absorption, these phenomena cannot carry a very large column density of neutral hydrogen nor contribute to the kinematics apparent in the cold flows unless they replace the streams altogether. This would fail to supply the gas necessary to produce the observed high SFR plus a similar outflow rate. Unless these SFRs and outflows are sustained by a large amount of reservoir gas, the inflows and outflows must live together. This is a natural consequence of the fact that the inflows are along narrow dense streams, while the outflows are covering a large solid angle between the streams. Additional work in simulations is needed to explore stronger wind models and to understand the interaction between outflowing and inflowing gas (e.g. Razoumov 2009, Powell et al. 2010), before we can reconcile both the Lyα and metal absorption profiles with a combination of inflows and outflows.

2.8 Summary and Conclusions

We investigated the characteristics of cold gas in cold-stream fed galaxies in absorption using a sample of seven galaxies drawn at random for zoom-in cosmological hydro-AMR simulations with high resolution ($35 - 70$ pc), in the halo mass range $10^{10} - 10^{12}$ $M_{\odot}$ between $z \sim 1.4 - 4$. We considered the contributions of the cold gas in the central galaxies and of the inflowing cold gas in the two components of the streams that feed the high-redshift massive galaxies from the cosmic web — the incoming small galaxies and the larger smooth component. Our analysis focused on the mass dependence and time evolution of optically thick and thin

gas cross section and covering factor, on the metallicity distribution in the streams and massive galaxies, and on the Lyα and metal line kinematics. The simulations were compared with available observations of absorption line systems in the foreground of quasars and with kinematics of the circumgalactic medium in Lyman break galaxies to highlight what observables are most useful to detect cold streams in absorption. The limited sample size and the fact that very strong outflows are not included in these simulations make our predictions rather specific to the study of cold streams, and caution is advised in generalizing our findings to the full population of ALSs. Our results and findings can be summarized as follows.

- After post processing the simulations with a radiative transfer code that includes dust, collisional ionization, and photoionization from the UV background and local stellar sources, cold streams appear highly ionized and not entirely self-shielded structures. Pockets of high column density neutral hydrogen, mostly associated with the central galaxy and in-streaming galaxies, are embedded in a ionized medium with $N_{\rm HI} \lesssim 10^{19}$ cm^{-2}. The densest regions are mostly affected by the UV radiation from local sources because the UV background is completely shielded in regions of $n_{\rm H} > 0.1$ cm^{-3}. While the UV from stars is largely confined by hydrogen and dust absorption in the star-forming regions, the non-zero escape fraction is responsible for the ionization of the gas in the immediate surroundings of disks ($10^{18} \lesssim N_{\rm HI}/{\rm cm}^{-2} \lesssim 10^{20}$). Since most of the mass is found at the highest column densities, UV from stars decreases the total neutral hydrogen mass in galaxies.

- The smooth component of the cold streams dominates the cross section of neutral hydrogen below $N_{\rm HI} = 10^{18}$ cm^{-2} within $2R_{\rm vir}$. At $z \leq 2$, less than half of the cross section in the range $N_{\rm HI} = 10^{19} - 10^{20}$ cm^{-2} is due to the streams. At these low redshifts, most of the cross section for $N_{\rm HI} > 10^{20}$ cm^{-2} is in galaxies, with the smooth component of the streams accounting only for a small fraction of the total budget. However, a rapid evolution is seen

with redshift, with the gas clumps in the streams responsible for more than half of the DLA/SLLS covering factor at $z \sim 4$.

- No more than 60% of the DLA cross section is associated with the central galaxies. Satellites and gas clumps in streams are important or even the dominant contributors to the neutral gas cross section, especially in the most massive galaxies.

- At all redshifts, the covering factor at $2R_{\rm vir}$ is $< 40\%$ for $N_{\rm HI} > 10^{15}$ cm^{-2}, $< 10\%$ for LLSs, and $< 1\%$ for DLAs. The H I gas covering factor within $2R_{\rm vir}$ appears to decrease slowly from $z \sim 4$ to $z \sim 1.5$, and the covering factor within a fixed aperture of 10 arcsec ($\sim 70 - 85$ kpc) seems not evolving or slowly increasing with redshift. However, the small sample and the non-homogeneous sampling of the halo mass function at all the redshifts prevent us from deriving robust scaling relations. In these simulations, a certain degree of self-similarity is found in the column density probability distribution function for neutral hydrogen and the time evolution of the different covering factors appears to follow the universal expansion.

- Cold streams mostly appear as LLSs and contribute significantly to the column density distribution function in between $N_{\rm HI} = 10^{17} - 10^{18}$ cm^{-2}. Massive disks, as described by these high resolution simulations, shape the column density distribution function for very high column density DLAs. These models, especially if molecules are included, can reproduce the observed knee at $10^{21.5}$ cm^{-2} in the $f(N_{\rm H\,I}, X)$. Conversely, there is a deficiency of systems between $10^{18.5} - 10^{20}$ cm^{-2}, likely due to the missing low mass systems in our simulated sample and to finite numerical resolution which does not capture a clumpy medium. The simulated $f(N_{\rm H\,I}, X)$ is very shallow below 10^{17} cm^{-2}, indicating that cold streams in proximity to massive galaxies are not the only sources of opacity for Lyman limit photons in the Universe; an additional contribution from the IGM is required.

Finally, in agreement with observations, the simulated $f(N_{\rm H\,I}, X)$ is not evolving in shape with redshift, due to a nearly self-similar distribution of neutral hydrogen in halos.

- Despite the small covering factor, comparing the observed and simulated incidence of absorption line systems, it appears that massive galaxies and cold streams are responsible for at least 30% of the absorbers detected in the foreground of QSOs. In fact, due to a large enough number density of galaxies in the universe, the cross section required to match the observed incidence is only a factor of a few above the mean values in this sample. A modest extrapolation towards lower halo masses can fully reproduce the observed population. To the extent that cold streams are indeed the common mode of accreting gas in massive galaxies at high redshift, as emerges from the cosmological simulations, these streams have already been detected in absorption line surveys, primarily as metal-poor LLSs.

- In our simulations, most of the cross section within $2R_{\rm vir}$ is occupied by metal poor gas ($Z \sim 10^{-2} Z_{\odot}$), that has been partly enriched by previous episodes of star formation. DLAs and SLLSs exhibit broad metallicity distributions, with a peak at $Z \sim 10^{-1} Z_{\odot}$ and $Z \sim 10^{-1.5} Z_{\odot}$, respectively. Metallicity is a valuable tool to disentangle gas that is associated with galaxies or winds (with metallicity distribution skewed above $Z \sim 10^{-1.5} Z_{\odot}$) from LLSs in the cold streams (with the lowest level of enrichment).

- Because of the partial covering factor of neutral gas, composites of Lyα absorption lines are not black at the line center, even within 40 kpc from the central galaxy. The typical line width outside the central galaxy and its immediate neighborhood is $\sim 200 - 300$ km/s FWHM reflecting the bulk inflow velocity in cold streams. Our current simulations reproduce the Lyα equivalent width distribution observed in the circumgalactic medium of Lyman break galaxies. Conversely, the predicted optical depth of metal lines is suppressed by more than 5 orders of magnitude and the predicted equivalent width of Si IIλ1260

is systematically lower than the observed values. An increase in the metallicity of the streams is not enough to match the observations. Another mechanism is required, such as massive outflows beyond the existing winds in the current simulations.

From the systematic analysis of our simulations and a careful comparison with observations, we conclude that the theoretical predictions of cold gas accretion along filaments of the cosmic web are consistent with present day observations. This conclusion is not driven by a null claim that the cold streams are not detectable in absorption, which would have been rather unsatisfactory. In fact, our current simulations indicate that cold streams contribute significantly to the incidence and kinematics observed at redshift $z \sim 2-3$.

Significant progress can be made in two directions. Observationally, it should be possible to identify candidates of cold streams by searching the increasing spectroscopic data sets of quasars and galaxy pairs. Despite the difficulties of measuring metallicity in LLSs, one can investigate their kinematics in detail to disentangle between metal rich gas that is outflowing at velocities > 600 km/s from gas that is inflowing at velocities $\sim< 200$ km/s with metallicity $\lesssim 0.01 Z_{\odot}$. At the same time, improved models of outflows in simulations are required to formulate complete predictions of the interactions between inflowing and outflowing gas that are crucial to guide observers in the interpretation of their results. Following these paths, it may become possible in the near future to uncover the modes of gas accretion in high redshift galaxies, and their interplay with outflows.

2.9 Appendix

2.9.1 A detailed comparison of the RT calculations

In this appendix, we provide a more detailed comparison of the results from the different radiative transfer calculations. At first, we discuss the effects that the UVB and local sources have on the column density distribution. Then, we quantify the hydrogen density threshold above which the gas is self-shielded and we provide a numerical approximation useful to correct the CIE approximation in the absence of the UVB. Finally, we examine the impact that photoionization has on the neutral mass of galaxies and satellites.

2.9.1.1 Effects on the hydrogen column density distribution

In Figure 2.17, we highlight the effect that different ionization sources have on the H I column density distribution by showing the number of AMR cells in bins of $N_{\rm HI}$ from an arbitrary projection of MW3 at $z = 2.3$. The different RT models have been normalize with respect to the STAR model (black horizontal solid line). The CIE model (blue dash-dotted line) offers an upper limit to the neutral fraction at all the column densities. The CIE approximation holds only in highly optically thick regions ($N_{\rm HI} \gtrsim 10^{20}$ cm^{-2}), while at lower column density overpredicts by a large amount the neutral fraction in comparison to the UVB case (red dashed line). Photoionization cannot be neglected, particularly given the fact that streams are not entirely self-shielded, as evident from Figure 2.3.

Comparing the UVB with the STAR model, it appears that radiation from local sources affects primarily the high column densities ($N_{\rm HI} \gtrsim 10^{21}$ cm^{-2}) which are completely self-shielded from the UVB. Since most of the ionizing photons are absorbed locally, the effect of local sources is modest below $N_{\rm HI} \sim 10^{20}$ cm^{-2} and almost negligible below $N_{\rm HI} \sim 10^{18}$ cm^{-2}. Note that quantities shown in Figure 2.17 are normalized and not proportional to the area covered by gas

at a given column density. In fact, radiation from local sources that leaks form the central disk can affect the H I column density distribution function in the interval $10^{18} - 10^{20}$ cm^{-2}, when weighted by area. Dust produces little or no effect to the UVB model (compare the red dashed line with open squares).

Although not included in the our RT calculation, in Figure 2.17 we explore how molecules affect the shape of the hydrogen column density distribution. When atomic hydrogen is shielded from UV radiation, molecules can form, lowering the effective H I column density. To investigate the importance of this effect, we compute the molecular fraction following the McKee & Krumholz (2010) formalism. This analytic model reproduces to first order the molecular gas fraction in nearby spirals (Krumholz et al. 2009a) and metal poor dwarfs (Fumagalli et al. 2010a) and approximates the results from numerical simulation (Krumholz & Gnedin 2011). In each AMR cell, we estimate the molecular gas fraction as a function of the total neutral gas column density $\Sigma_{\rm gas} = 2m_p N_{\rm H_2} + m_p N_{\rm HI}$ and metallicity. Since this model is designed to describe the molecular fraction in individual clouds, we correct for smearing due to resolution by assuming $\Sigma_{\rm cloud} = c\Sigma_{\rm gas}$ with $c = 5$ for all the AMR cells with sizes above 200 pc and $c = 1$ elsewhere. Although not a dominant correction, the inclusion of molecules lowers the H I column densities above $N_{\rm HI} \sim 10^{22}$ cm^{-2} (open circles). For this reason, the values of $N_{\rm HI}$ used in our analysis should be regarded as upper limits at the highest density. This column density threshold is in good agreement with the knee in the observed DLA column density distribution (Prochaska et al. 2005, Noterdaeme et al. 2009) that, indeed, is interpreted as the transition between atomic and molecular gas (e.g. Zwaan & Prochaska 2006).

2.9.1.2 Self shielding and numerical approximations

Figure 2.18 shows the neutral fraction $x_{\rm HI}$ as a function of total gas volume density $n_{\rm H}$, for the CIE model (top left panel), the UVB model (top right panel), and the STAR model

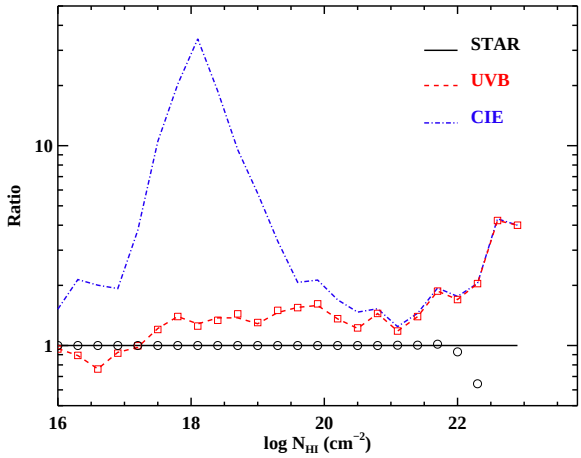

Figure 2.17: Comparison of the neutral hydrogen column density in the STAR model (black line), the CIE model (blue dash-dotted line), and the UVB model, with and without dust (red dashed line and open squares). The STAR model corrected for molecular hydrogen is show with open circles. For each RT calculation, we plot the number of projected cells in a given interval of $N_{\rm HI}$, in comparison to the STAR model. The CIE approximation holds only in highly optically thick regions. At high column densities, neutral hydrogen is further ionized by local sources and depleted by molecules.

(bottom left panel), again for MW3 at $z = 2.3$. In presence of the UVB, hydrogen is highly ionized below $n_{\rm H} = 0.01$ cm^{-3}, whereas it is fully neutral above $n_{\rm H} = 0.1$ cm^{-3}. A similar volume density for self-shielding has been consistently found with different numerical techniques (e.g. Razoumov et al. 2006, Faucher-Giguère et al. 2010), justifying the use of a fixed threshold above which the UVB can be suppressed while running the simulations.

Additional photo-heating, currently neglected in our RT calculations, will produce a smoother transition between ionized and shielded regions due the enhanced collisional ionization in partially shielded cells. Also, both the UVB and CIE models do not account for the presence of ionizing sources that increase the dispersion of $x_{\rm HI}$ (bottom left panel in Figure 2.18). A significant fraction of the gas above $n_{\rm H} = 0.01 - 0.1$ cm^{-3} remains fully neutral, but local fluctuations in the ionization state of up to 5 order of magnitude can be found. Indeed, a locus of points with $n_{\rm H} > 0.01$ cm^{-3} and $x_{\rm HI} < 0.01$ is visible in Figure 2.18. Since this gas mostly

resides in the central disk and in satellites where the majority of the star formation is present, the inclusion of ionizing photons seem to be an essential ingredient in models of galaxy formation. Ionizing photons from stars are absorbed mostly in the inner part of the galaxy by dust and gas and there is only a minor contribution of the ionization below $n_{\rm H} = 0.01$ cm^{-3}, in qualitative agreement with current constraints on the escape fraction of ionizing radiation. However, the degree of precision required in comparing models and observations requires the inclusion of this RT effects on the neutral gas, commonly neglected or treated with some approximations.

Although below $n_{\rm H} = 0.01$ cm^{-3} the scatter in $x_{\rm HI}$ at a given volume density is large, most of the AMR cells accumulate in a well defined locus of points. Due to a self-similarity in the probability distribution function of the hydrogen column density, there is little scatter for different galaxies both with mass and redshift, as seen from the bottom right panel of Figure 2.18, where we plot the arithmetic mean of $x_{\rm HI}$ for the UVB model (gray lines for individual galaxies). This invariance not only corroborates the assumption of a fixed threshold above which the UVB is suppressed, but also allows one to assume a lower limit on the hydrogen volume density in systems that host neutral gas, such as DLAs. Further, we can derive a crude approximation to the UVB model. In the CIE model, fully neutral or marginally ionized gas is found even below $n_{\rm H} = 0.01$ cm^{-3}, with an error up to several order of magnitudes in comparison to the UVB model. Using the fact that the mean $x_{\rm HI}$ as a function of $n_{\rm H}$ is sensitive to the intensity of the UV background only, we derive a relation $x_{\rm HI} = x_{\rm HI}(n_{\rm H})$ that improves the CIE model in presence of a UVB. We obtain such relation by combining all the arithmetic means computed in individual galaxies with a geometric mean (see Figure 2.18 and Table 2.3).

2.9.1.3 Neutral hydrogen mass

For each RT model, we measure the total H I mass enclosed in the virial radius of the main halos and satellites (see Table 2.2). A comparison is provided in Figure 2.19. In the

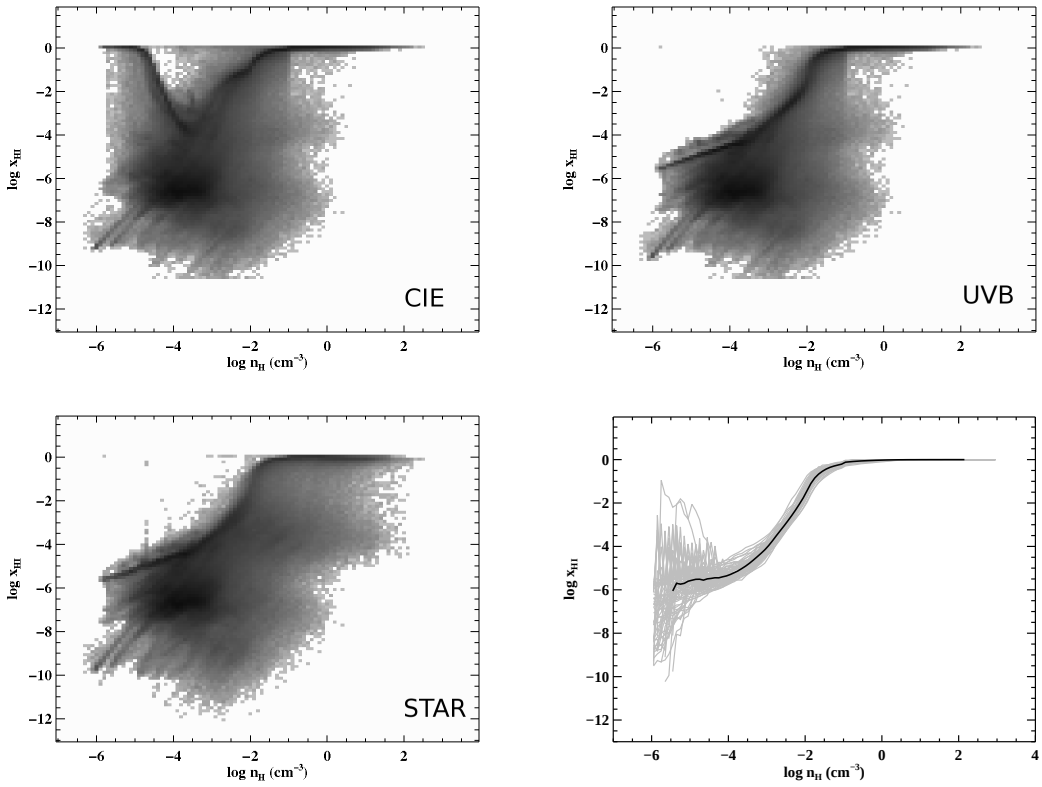

Figure 2.18: Neutral fraction as a function of the hydrogen volume density for the CIE model (top left), the UVB model (top right) and the STAR model (bottom left) in MW3 at $z \sim 2.3$. The color scale is proportional to the number of AMR cells in each bin, with the most populated region in dark gray. The arithmetic mean in bin of $\log n_{\rm H}$ for the UVB model in individual galaxies is shown in the bottom right panel (gray lines). The central values common to all the galaxies is also superimposed (black solid line). Although with significant scatter, this distribution can be used as a crude approximation to the UVB model. RT with local sources is required to capture the intrinsic scatter in the neutral fraction above the $n_{\rm H} \sim 0.01 - 0.1$ cm^{-3}, where gas becomes shielded from the UVB.

top panel, we show the relation between the total and the neutral hydrogen mass in the STAR model. In massive halos above 10^9 $M_\odot$, less than $\sim 1/5$ of the hydrogen mass is neutral. A robust linear regression in logarithmic space for the main halos only gives $\log M_{\rm HI} = 0.03 + 0.93 \log M_{\rm H}$. Less massive satellites cover a wider range in H I mass, from almost neutral to highly ionized. The CIE and UVB models predict nearly identical masses above 10^9 $M_\odot$ ($\log M_{\rm HI,CIE} = 0.02 + 1.00 \log M_{\rm HI,UVB}$ for the main halos), while in the satellites the CIE model predicts larger neutral hydrogen masses. H I masses from the STAR model are systematically lower than those in the CIE or the UVB model above 10^9 $M_\odot$, up to 50%-60% above 10^{10} $M_\odot$ ($\log M_{\rm HI,STAR} = 0.41 + 0.93 \log M_{\rm HI,CIE}$ and $\log M_{\rm HI,STAR} = 0.45 + 0.93 \log M_{\rm HI,UVB}$ for the main halos). Satellites have a larger scatter.

The explanation for this behavior naturally emerges from the previous considerations. In the main halos, most of the mass is in high density regions where the gas is self-shielded from the UVB. Photoionization from local sources and, to second order, electron collisions regulate the amount of mass in neutral hydrogen. For satellites, the observed scatter is due to the different SFRs in these systems and on the different distances from the central disks. Also, at lower masses, systems are more vulnerable to ionization from the UVB. This behavior is qualitatively in agreement with the simulations by Gnedin (2010) who finds that the Lyman-Werner radiation from the UVB is negligible compared to the local radiation field and that the ionizing radiation from the UVB plays a significant role only up to the edge of the H I disks. This result has implications for understanding the processes that shape the massive end of the H I mass function and stresses further the necessity to account for local photoionization as an important feedback mechanism (e.g. Cantalupo 2010).

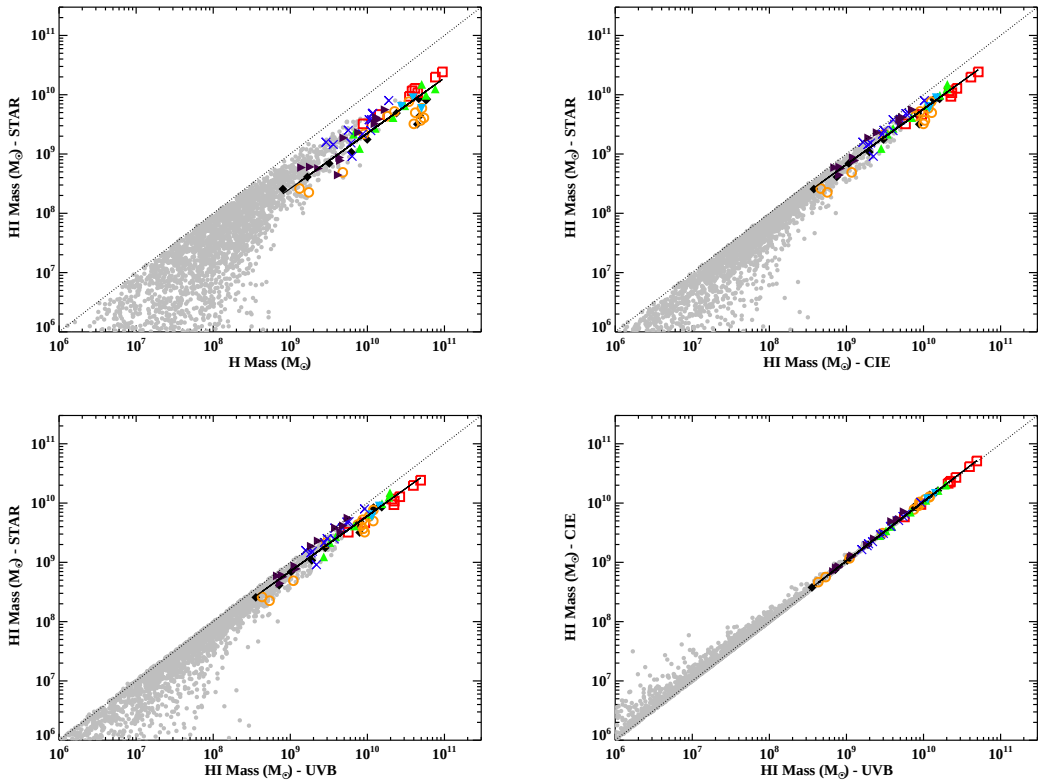

Figure 2.19: Comparison of the neutral and total hydrogen mass enclosed within $R_{\rm vir}$ for the three RT models. In each panel, main halos are color coded as in Figure 2.1. Satellites are displayed with gray circles. Dotted lines indicate a 1:1 relation, while dashed lines show a robust linear regressions for the main halos only. In massive galaxies, only 20-25% of the total hydrogen mass is neutral, with local sources and electron collisions being responsible for most of the ionization. The UVB has no effects on the neutral mass due to nearly complete self-shielding at the highest hydrogen volume densities.

2.9.1.4 Uncertainties and future improvements

Although our model significantly improves upon simpler RT calculations, few approximations contribute the final error budget. For example, these simulations do not include the presence of an active galactic nucleus and additional ionization from this harder spectrum is neglected in this work. Similarly, we include only photoionization at 912Å due to stars and we do not considering harder SEDs that could enhance the ionization via heating. Similarly, He recombination emission is neglected. Finally, we ignore the possibility that these galaxies could lie in the proximity of other ionizing sources, such as QSOs (e.g. Cantalupo et al. 2005). We note the these effects would reduce further the neutral gas fraction and enhance the effects of ionizing radiation on the neutral volume and column density.

In the previous sections, we have clearly shown that the inclusion of both the UVB and local sources shapes the distribution of gas volume and column density at the low and high densities. However, the exact amplitude of these effects is proportional to the intensity of the input radiation field which is subject to uncertainties, as discussed in Section 2.4.1. Heating from the UVB and local sources is already included in these simulations, but during the post-processing we do not update the gas temperature (see a discussion in Cantalupo & Porciani 2011). And a sensible determination of the temperature has obvious implications for collision ionization and emission properties (Faucher-Giguère et al. 2010).

We try to quantify the impact that an increase in temperature has on our results by performing a simple test with MW3 at $z \sim 2.3$, using the STAR model. After having increased by 50% the temperature in photo-ionized cells with $x_{\rm HI} < 0.8$, we recompute the neutral fraction for CIE ($x_{\rm HI,cie}$). Then, we derive the projected H I distribution function using an updated neutral fraction $x_{\rm HI,new} = \min(\sqrt{x_{\rm HI,old} x_{\rm HI,cie}}; x_{\rm HI,old}; x_{\rm HI,cie})$, defined in this way to favor high ionization. Results are summarized in Figure 2.20. In the bottom panel, we show the column

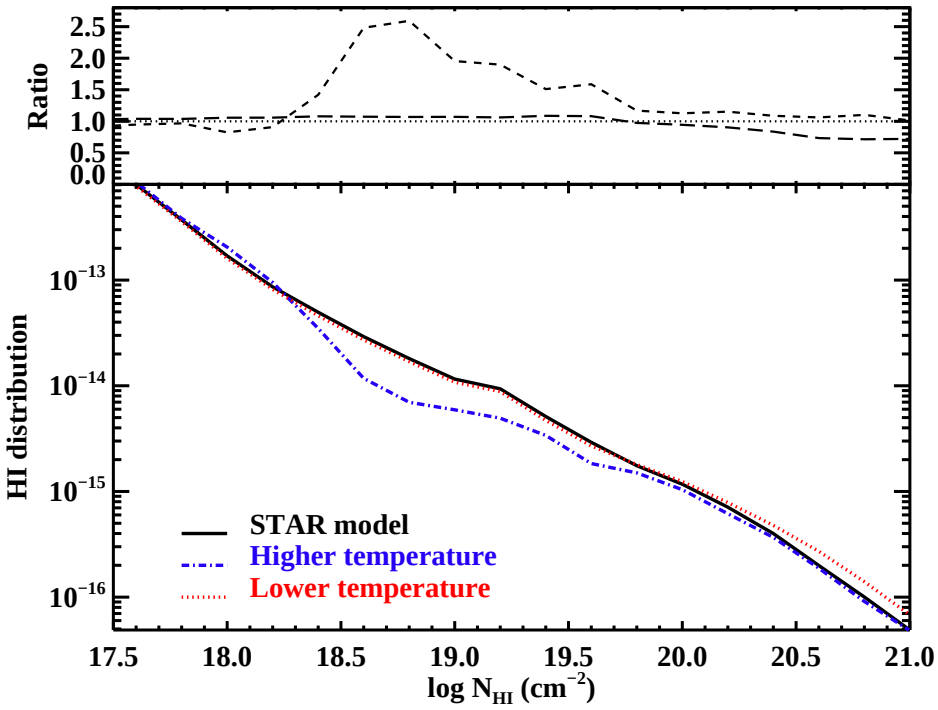

Figure 2.20: Bottom panel: column density distribution function for the STAR model (black solid line), for a model with increased gas temperature (blue dash line) and a model at lower temperature (red dotted line). Top panel: the ratio of the distribution for increased (decreased) temperature and the STAR model is displayed with a dashed (long dashed) line. Identity between the models is marked with a dotted line. Additional heating produces higher ionization in the interval $\log N_{\rm HI} = 18.3 - 19.8$ cm^{-2}, but the covering factor of optically thick gas is nearly unperturbed, owing to a rising $f(N_{\rm H\,I}, X)$ towards lower column densities. Lowering the temperature, little differences (less than a factor of two) are visible since the UV radiation is the dominant source of ionization in unshielded regions, while shielded regions are already cold and neutral.

density distribution function for the original STAR model (black solid line) and for the model at higher temperature (blue dash line). In the top panel we display the ratio of the two distributions. Additional heating produces higher ionization in the interval $\log N_{\rm HI} = 18.3{-}19.8$ cm^{-2}, yielding a flatter $f(N_{\rm H\,I}, X)$. The cross section of SLLSs is reduced by a factor of ~ 2, increasing even further the discrepancy with observations. The DLA cross section is decreased by no more than 20% and the effect on the cumulative covering factor of LLSs is small, owing to a rising $f(N_{\rm H\,I}, X)$ towards lower column densities. This simple test suggests that a moderate increase is temperature does not have a major impact on the result presented in this paper. Conversely, if cooling is more effective than what assumed in these simulations, hydrogen is less ionized by electron collisions. In Figure 2.20, we shown the hydrogen distribution (red dotted line) for a model in which the temperature is reduced to 50% of its value in all the cells with $n_{\rm H} > 0.05$ cm^{-2}, i.e. partially or totally shielded. Gas below this density is highly photoionized and the ionization state is insensitive to the gas temperature. In this case, the ionization fraction is assumed to be $x_{\rm HI,new} = x_{\rm HI,old}$ for $n_{\rm H} < 0.05$ cm^{-2} and $x_{\rm HI,new} = x_{\rm HI,cie}$ for $n_{\rm H} \geq 0.05$ cm^{-2}. Since gas above $n_{\rm H} \sim 0.05$ cm^{-2} is already at low temperature, little difference (less than a factor of two) is visible in the final H I distribution.

2.9.2 Additional Tables

Table 2.1: Summary of the baryonic and dark matter properties of the 7 galaxies used in this study.

Redshift	MW1	MW2	MW3	MW4	MW5	MW8	MW9
				Virial Mass (10^{11} M$_\odot$)			
4.00	0.11	0.91	0.29	1.26	6.09	0.62	0.19
3.55	0.26	1.51	0.37	1.73	9.58	0.73	0.33
3.17	0.53	3.14	1.06	2.73	11.70	1.13	0.48
2.85	1.13	4.19	1.62	4.23	-	1.52	1.22
2.57	1.68	5.04	2.66	5.81	-	1.80	1.31
2.33	3.34	5.92	3.55	9.21	-	2.31	1.29
2.13	6.49	7.64	5.73	12.01	-	2.54	1.32

Continued on Next Page...

Table 2.1 – Continued

Redshift	MW1	MW2	MW3	MW4	MW5	MW8	MW9
1.94	8.83	9.12	8.06	15.17	-	2.79	1.75
1.78	10.20	-	9.27	-	-	2.85	2.87
1.63	11.47	-	11.81	-	-	2.85	3.00
1.50	13.03	-	12.59	-	-	2.86	3.39
1.38	13.69	-	13.52	-	-	5.10	4.31
			Virial Radius (kpc)				
4.00	15.00	30.00	20.50	33.50	56.50	26.25	18.25
3.55	21.75	39.00	24.50	41.00	72.25	30.50	23.50
3.17	29.75	57.00	38.00	51.75	84.25	38.50	29.00
2.85	41.75	64.50	47.25	64.50	-	46.00	42.75
2.57	51.50	74.25	60.50	77.25	-	52.25	47.00
2.33	68.75	83.25	70.50	96.50	-	60.75	50.00
2.13	91.25	97.25	88.00	112.25	-	66.75	53.50
1.94	106.75	108.75	104.00	129.00	-	73.00	62.50
1.78	118.50	-	114.75	-	-	77.75	77.50
1.63	129.50	-	131.25	-	-	81.75	82.75
1.50	141.75	-	140.25	-	-	85.50	90.50
1.38	150.50	-	150.00	-	-	108.00	102.25
			Star Formation Rate ($M_\odot$ yr^{-1})				
4.00	0.94	37.79	3.74	21.19	49.49	4.23	2.02
3.55	4.61	58.25	2.70	28.66	140.61	5.70	2.88
3.17	4.42	148.23	14.44	15.77	76.14	4.48	3.12
2.85	14.76	157.48	5.54	38.11	-	40.65	6.98
2.57	9.65	104.36	25.95	48.22	-	4.73	8.28
2.33	35.71	99.77	30.49	44.21	-	7.16	7.14
2.13	52.47	143.23	46.77	73.75	-	5.03	2.45
1.94	62.03	187.86	71.46	77.22	-	13.63	2.38
1.78	108.52	-	84.00	-	-	6.26	5.44
1.63	20.71	-	50.19	-	-	3.42	17.68
1.50	25.65	-	39.10	-	-	3.61	4.25
1.38	23.19	-	45.64	-	-	5.31	6.77
			Stellar Mass (10^{10} $M_\odot$)				
4.00	0.03	1.62	0.20	1.22	6.25	0.63	0.13
3.55	0.17	3.12	0.26	1.69	10.66	0.80	0.28
3.17	0.31	7.52	0.86	2.79	13.32	1.31	0.41
2.85	0.87	11.16	1.14	3.91	-	1.70	1.27
2.57	1.14	14.24	1.76	4.94	-	1.85	1.48
2.33	2.25	17.51	2.63	8.69	-	2.21	1.61
2.13	5.54	22.26	4.34	12.07	-	2.42	1.70
1.94	8.19	27.48	7.14	15.38	-	2.96	1.94
1.78	10.57	-	8.90	-	-	3.04	3.19
1.63	11.43	-	12.10	-	-	3.09	3.46
1.50	12.80	-	13.15	-	-	3.20	3.98
1.38	14.19	-	14.38	-	-	6.42	5.49
			Dark Matter Mass (10^{11} $M_\odot$)				

Continued on Next Page...

Table 2.1 – Continued

Redshift	MW1	MW2	MW3	MW4	MW5	MW8	MW9
4.00	0.10	0.66	0.25	1.07	5.19	0.53	0.16
3.55	0.23	1.06	0.33	1.49	8.13	0.61	0.29
3.17	0.46	2.04	0.93	2.32	9.86	0.95	0.41
2.85	0.98	2.66	1.42	3.62	-	1.29	1.05
2.57	1.46	3.24	2.30	5.01	-	1.55	1.11
2.33	2.88	3.72	3.07	7.84	-	2.01	1.09
2.13	5.54	4.65	4.95	10.23	-	2.20	1.11
1.94	7.54	5.43	6.92	12.88	-	2.39	1.48
1.78	8.70	-	7.98	-	-	2.43	2.42
1.63	9.85	-	10.11	-	-	2.42	2.53
1.50	11.17	-	10.77	-	-	2.42	2.86
1.38	11.68	-	11.54	-	-	4.27	3.59
			Hydrogen Mass (10^{10} $M_\odot$)				
4.00	0.080	0.879	0.132	0.663	2.761	0.291	0.140
3.55	0.167	1.431	0.174	0.791	3.921	0.361	0.182
3.17	0.320	3.517	0.480	1.265	5.041	0.566	0.232
2.85	0.619	4.138	0.906	2.156	-	0.635	0.494
2.57	0.996	3.803	1.823	3.057	-	0.614	0.447
2.33	2.390	4.502	2.249	5.065	-	0.862	0.419
2.13	3.990	7.622	3.453	5.778	-	1.058	0.437
1.94	4.648	9.399	4.181	7.590	-	1.109	0.767
1.78	4.449	-	4.009	-	-	1.136	1.312
1.63	4.791	-	4.926	-	-	1.162	1.255
1.50	5.763	-	5.097	-	-	1.177	1.382
1.38	5.920	-	5.413	-	-	1.895	1.701

Table 2.2: Summary of the neutral gas properties of the 7 galaxies used in this study.

Redshift	MW1	MW2	MW3	MW4	MW5	MW8	MW9
			H I Mass (10^{10} $M_\odot$) - UVB				
4.00	0.036	0.565	0.043	0.326	1.124	0.158	0.068
3.55	0.072	0.926	0.054	0.271	1.423	0.186	0.079
3.17	0.103	2.210	0.110	0.389	1.147	0.302	0.081
2.85	0.190	2.631	0.295	0.676	-	0.218	0.187
2.57	0.285	2.085	0.751	1.071	-	0.194	0.120
2.33	0.805	2.230	0.869	1.963	-	0.277	0.074
2.12	1.454	3.950	1.195	1.548	-	0.379	0.115
1.94	1.537	4.902	1.193	1.902	-	0.372	0.230
1.78	0.794	-	0.922	-	-	0.476	0.488
1.63	0.745	-	0.885	-	-	0.544	0.452
1.50	1.207	-	1.048	-	-	0.579	0.385

Continued on Next Page...

Table 2.2 – Continued

Redshift	MW1	MW2	MW3	MW4	MW5	MW8	MW9
1.38	1.202	-	0.818	-	-	0.922	0.560
H I Mass (10^{10} $M_\odot$) - STAR							
4.00	0.026	0.322	0.026	0.215	0.637	0.158	0.059
3.55	0.041	0.464	0.023	0.124	0.891	0.145	0.060
3.17	0.069	0.940	0.049	0.270	0.589	0.252	0.057
2.85	0.108	1.281	0.217	0.410	-	0.091	0.187
2.57	0.176	1.169	0.434	0.651	-	0.153	0.078
2.33	0.490	1.081	0.522	1.494	-	0.216	0.044
2.12	0.901	1.983	0.763	0.977	-	0.379	0.087
1.94	0.850	2.427	0.498	1.252	-	0.248	0.230
1.78	0.318	-	0.323	-	-	0.379	0.415
1.63	0.441	-	0.369	-	-	0.478	0.324
1.50	0.810	-	0.591	-	-	0.487	0.382
1.38	0.839	-	0.407	-	-	0.798	0.559

Table 2.3: Geometric mean of the $x_{\rm HI}$ vs $n_{\rm H}$ relation from the UVB model.

$\log n_{\rm H}$ ($\rm cm^{-3}$)	$\log x_{\rm HI}$	$\log n_{\rm H}$ ($\rm cm^{-3}$)	$\log x_{\rm HI}$
-4.750	-5.512E+00	-2.950	-3.959E+00
-4.650	-5.554E+00	-2.850	-3.739E+00
-4.550	-5.496E+00	-2.750	-3.506E+00
-4.450	-5.475E+00	-2.650	-3.279E+00
-4.350	-5.445E+00	-2.550	-3.054E+00
-4.250	-5.448E+00	-2.450	-2.820E+00
-4.150	-5.401E+00	-2.350	-2.579E+00
-4.050	-5.360E+00	-2.250	-2.323E+00
-3.950	-5.293E+00	-2.150	-2.050E+00
-3.850	-5.212E+00	-2.050	-1.764E+00
-3.750	-5.128E+00	-1.950	-1.432E+00
-3.650	-5.020E+00	-1.850	-1.098E+00
-3.550	-4.906E+00	-1.750	-8.472E-01
-3.450	-4.775E+00	-1.650	-6.688E-01
-3.350	-4.631E+00	-1.550	-5.363E-01
-3.250	-4.479E+00	-1.450	-4.328E-01
-3.150	-4.321E+00	-1.350	-3.546E-01
-3.050	-4.155E+00	-1.250	-2.978E-01
-1.150	-2.522E-01	0.550	-6.192E-03
-1.050	-2.097E-01	0.650	-5.050E-03
-0.950	-1.192E-01	0.750	-4.217E-03
-0.850	-1.012E-01	0.850	-3.391E-03
-0.750	-8.607E-02	0.950	-2.770E-03
-0.650	-7.321E-02	1.050	-2.307E-03

Continued on Next Page...

Table 2.3 – Continued

$\log n_{\rm H}$ ($\rm cm^{-3}$)	$\log x_{\rm HI}$	$\log n_{\rm H}$ ($\rm cm^{-3}$)	$\log x_{\rm HI}$
-0.550	-6.224E-02	1.150	-1.850E-03
-0.450	-5.388E-02	1.250	-1.530E-03
-0.350	-4.524E-02	1.350	-1.256E-03
-0.250	-3.801E-02	1.450	-9.829E-04
-0.150	-3.123E-02	1.550	-6.561E-04
-0.050	-2.520E-02	1.650	-4.945E-04
0.050	-2.037E-02	1.750	-3.221E-04
0.150	-1.579E-02	1.850	-2.372E-04
0.250	-1.247E-02	1.950	-1.611E-04
0.350	-1.015E-02	2.050	-1.826E-04
0.450	-7.837E-03	2.150	-8.635E-05

Table 2.4: DLAs and LLSs covering factor from the STAR model. The top table lists values for the streams and galaxies, while the bottom one lists values for the streams alone.

Redshift	MW1	MW2	MW3	MW4	MW5	MW8	MW9
	DLA covering factor in % at $2R_{\rm vir}$ (streams and galaxies)						
4.00	1.78	3.47	1.09	2.68	1.42	1.57	1.55
3.55	1.03	3.65	0.79	1.35	1.22	0.85	1.02
3.17	0.94	1.95	0.60	1.76	1.08	0.88	0.76
2.85	0.79	2.65	1.42	1.86	-	0.28	0.62
2.57	1.12	2.84	0.88	1.60	-	0.31	0.29
2.33	1.37	2.22	1.19	1.54	-	0.31	0.49
2.12	1.07	2.23	0.75	1.00	-	0.59	0.82
1.94	0.62	2.07	0.38	0.72	-	0.28	0.73
1.78	0.23	-	0.17	-	-	0.25	0.66
1.63	0.36	-	0.21	-	-	0.51	0.30
1.50	0.35	-	0.27	-	-	0.51	0.35
1.38	0.26	-	0.32	-	-	0.48	0.44
	LLS covering factor in % at $2R_{\rm vir}$ (streams and galaxies)						
4.00	10.80	12.54	6.70	12.19	12.43	11.20	11.72
3.55	7.92	9.02	6.24	8.56	11.11	9.35	8.27
3.17	8.03	5.08	2.35	13.12	8.68	4.71	6.41
2.85	4.88	5.81	10.41	11.03	-	1.69	5.36
2.57	10.25	7.13	6.52	10.30	-	4.08	2.98
2.33	9.87	6.89	8.92	9.11	-	3.80	4.72
2.12	6.64	6.79	5.74	7.03	-	4.03	5.06
1.94	4.91	6.85	3.19	4.85	-	2.03	5.09
1.78	1.98	-	1.83	-	-	1.63	4.20
1.63	4.25	-	2.84	-	-	2.46	2.59
1.50	2.97	-	2.82	-	-	2.87	5.04

Continued on Next Page...

Table 2.4 – Continued

Redshift	MW1	MW2	MW3	MW4	MW5	MW8	MW9
1.38	1.84	-	2.40	-	-	1.66	4.21
DLA covering factor in % at $2R_{\rm vir}$ (streams only)							
4.00	1.47	2.94	0.82	2.41	0.77	0.95	1.29
3.55	0.70	3.02	0.52	0.98	0.68	0.34	0.74
3.17	0.62	1.48	0.36	1.28	0.39	0.33	0.57
2.85	0.41	2.24	1.00	1.27	-	0.05	0.37
2.57	0.69	2.40	0.32	0.91	-	0.03	0.15
2.33	0.70	1.89	0.54	0.95	-	0.08	0.15
2.12	0.50	1.83	0.32	0.50	-	0.15	0.26
1.94	0.22	1.75	0.16	0.44	-	0.03	0.37
1.78	0.07	-	0.04	-	-	0.03	0.34
1.63	0.08	-	0.07	-	-	0.24	0.05
1.50	0.07	-	0.07	-	-	0.15	0.12
1.38	0.02	-	0.14	-	-	0.28	0.24
LLS covering factor in % at $2R_{\rm vir}$ (streams only)							
4.00	10.55	12.05	6.30	11.63	11.82	10.60	11.42
3.55	7.46	8.28	5.78	7.95	10.41	8.75	7.83
3.17	7.51	4.41	1.87	12.45	7.85	3.96	6.09
2.85	4.17	5.23	9.67	10.24	-	1.13	4.85
2.57	9.73	6.58	5.51	9.34	-	3.40	2.49
2.33	8.99	6.42	7.88	8.29	-	3.13	4.04
2.12	5.66	6.31	5.10	6.26	-	3.31	4.21
1.94	4.23	6.41	2.60	4.38	-	1.56	4.50
1.78	1.62	-	1.44	-	-	1.02	3.42
1.63	3.64	-	2.47	-	-	1.98	2.04
1.50	2.30	-	2.31	-	-	2.22	4.30
1.38	1.34	-	1.97	-	-	1.19	3.64

Table 2.5: $f(N_{\rm H\,I}, X)$ from the STAR model witin $R_{\rm vir}$, $2R_{\rm vir}$ and in the streams alone.

log $N_{\rm HI}$ (cm^{-2})	$R_{\rm vir}$	$2R_{\rm vir}$	Streams
13.5	2.22×10^{-15}	8.18×10^{-15}	8.30×10^{-15}
14.5	2.61×10^{-16}	1.73×10^{-15}	1.74×10^{-15}
15.5	2.16×10^{-17}	1.35×10^{-16}	1.36×10^{-16}
16.5	1.94×10^{-18}	5.98×10^{-18}	6.06×10^{-18}
17.5	9.10×10^{-20}	3.03×10^{-19}	3.00×10^{-19}
18.5	2.91×10^{-21}	9.72×10^{-21}	9.52×10^{-21}
19.5	2.64×10^{-22}	7.70×10^{-22}	7.58×10^{-22}
20.5	2.65×10^{-23}	6.43×10^{-23}	5.46×10^{-23}
21.5	1.38×10^{-24}	2.55×10^{-24}	1.17×10^{-24}
22.5	7.80×10^{-27}	1.26×10^{-26}	1.78×10^{-28}

Continued on Next Page...

Table 2.5 – Continued

log $N_{\rm HI}$ (cm^{-2})	$R_{\rm vir}$	$2R_{\rm vir}$	Streams
23.5	9.92×10^{-30}	1.24×10^{-29}	–
24.5	4.08×10^{-32}	4.08×10^{-32}	–

Table 2.6: Average DLA cross sections within $R_{\rm vir}$ for all the halos at all redshifts.

Halo Mass ($M_\odot$)	Total cross section (kpc^2)		Central galaxy only (kpc^2)	
	Mean	Standard deviation	Mean	Standard deviation
3.6×10^{10}	46	24	30	21
8.6×10^{10}	138	93	82	58
1.6×10^{11}	209	170	104	86
2.6×10^{11}	312	225	182	145
3.9×10^{11}	723	396	251	184
6.8×10^{11}	623	498	215	252
1.1×10^{12}	522	252	216	163

Table 2.7: Equivalent widths for stacked Lyα and Si IIλ1260 absorption profiles at different impact parameters.

Radius (kpc)	W(Lyα) $z \sim 2.3$ (Å)	W(Si II) $z \sim 2.3$ (Å)	W(Lyα) $z \sim 3.2$ (Å)	W(Si II) $z \sim 3.2$ (Å)	W(Si II) enriched (Å)
0-10	7.753	0.295	6.454	0.200	0.329
10-41	2.053	0.050	1.300	0.023	0.082
41-82	0.875	0.010	0.578	0.003	0.023
82-125	0.679	0.006	0.390	0.001	0.014
125-200	0.465	0.003	-	-	0.006

Chapter 3

Detection of pristine gas two billion years after the Big Bang

This chapter reproduces the paper "Detection of pristine gas two billion years after the Big Bang" by Fumagalli et al. (2011b). Theoretical work, including our own (Fumagalli et al. 2011a), suggests how the study of Lyman limit systems offers a promising way to uncover the predicted mode of gas accretion onto galaxies. In this paper, we present the discovery of two Lyman limit systems at $z \sim 3$ and $z \sim 3.5$ that, despite their high hydrogen column density, do not show any appreciable metal content. Besides the relevance of this discovery in the context of Big Bang nucleosynthesis, this paper highlights how the physical properties of these two systems closely resemble the predictions of numerical simulations. Thus, these absorbers are ideal candidates for the long sought mode of gas accretion via cold streams. This work offers a proof-of-concept that large surveys of Lyman limit systems will be able to robustly test current theories of galaxy formation.

3.1 Abstract

In the current cosmological model, only the three lightest elements were created in the first few minutes after the Big Bang; all other elements were produced later in stars. To date, however, heavy elements have been observed in all astrophysical environments. We report the detection of two gas clouds with no discernible elements heavier than hydrogen. These systems exhibit the lowest heavy-element abundance in the early universe and thus are potential fuel for the most metal poor halo stars. The detection of deuterium in one system at the level predicted by primordial nucleosynthesis provides a direct confirmation of the standard cosmological model. The composition of these clouds further implies that the transport of heavy elements from galaxies to their surroundings is highly inhomogeneous.

3.2 Analysis and results

In modern cosmological theory, the light elements and their isotopes are produced during the first few minutes after the Big Bang when the universe cools during expansion from temperatures $T \sim 10^9\,\mathrm{K}$ to below $\sim 4 \times 10^8\,\mathrm{K}$. In this brief epoch, termed Big Bang Nucleosynthesis (BBN), D, ^{3}He, ^{4}He, and ^{7}Li are synthesized with an abundance ratio relative to hydrogen that is sensitive to the cosmic density of ordinary matter (i.e. the baryon density $\Omega_{b,0}$). BBN theory also predicts negligible production of the heavy elements with abundance ratios $\mathrm{X/H} < 10^{-10}$ and one must await the physical conditions that are typical of the stellar interiors (Burbidge et al. 1957).

The analysis of gas observed in absorption along the lines-of-sight to high-redshift quasars, distant galaxies that host supermassive black holes, is a powerful probe of the BBN yields. Particular attention has been given to deuterium, partly due to observational convenience but also because the D/H abundance ratio is very sensitive to $\Omega_{b,0}$. For quasar sight lines,

the measured $\log(\mathrm{D/H}) = -4.55 \pm 0.03$ (O'Meara et al. 2006, Pettini et al. 2008b) translates into $\Omega_{b,0}h^2(\mathrm{BBN}) = 0.0213 \pm 0.0010$, which is fully consistent with the value inferred from the Cosmic Microwave Background (CMB) power spectrum $\Omega_{b,0}h^2(\mathrm{CMB}) = 0.02249^{+0.00056}_{-0.00057}$ (WMAP7 Komatsu et al. 2011). This excellent agreement between two essentially independent experiments stands as a marked triumph of the Big Bang theory.

On the other hand, all of the systems with measured D have heavy-element abundances that exceed, by many orders of magnitude, the BBN prediction. In fact, despite measurement of thousands of galaxies from the early universe (e.g. Prochaska et al. 2003, Erb et al. 2006) and careful study of the diffuse gas that permeates the universe (e.g. Schaye et al. 2003, Simcoe et al. 2004), one has yet to detect anything near primordial enrichment. For structures denser than the mean cosmic density (Figure 3.1), the high-redshift universe has exhibited a floor in the metallicity Z, the mass fraction of elements heavier than hydrogen, at $\sim 1/1000$ of the solar abundance ($Z_\odot$). Similarly, among several old and iron-poor stars, only one has a metallicity $Z \sim 10^{-4} Z_\odot$ (Caffau et al. 2011), with all the remaining having enhanced C or O abundances. The existence of a minimum level of enrichment at about 1/1000 solar has been associated with the metal production in Population III (PopIII) stars, primordial stars that form in metal free environments via H_2 cooling. In fact, models and numerical simulations (e.g. Mackey et al. 2003, Wise & Abel 2008) show that ejecta from this first stellar population can enrich the interstellar medium (ISM) of the host halos up to $\sim 10^{-3} Z_\odot$ and pollute the surrounding intergalactic medium (IGM) as soon as 1 billion years after the Big Bang ($z \sim 6$).

In this work, we report on the hydrogen and metal properties of two gas 'clouds' at $z \sim 3$, when the universe was only two billion years old. We observed quasars SDSS J113418.96+574204.6 ($z_{\mathrm{em}} = 3.522$) and Q0956 + 122 ($z_{\mathrm{em}} = 3.297$) on UT January 3 and 5, 2006 and UT April 7, 2006 with the HIRES spectrometer on the Keck I telescope on Mauna Kea (see SOM text 1). Previous low-resolution spectra of these quasars from the Sloan Dig-

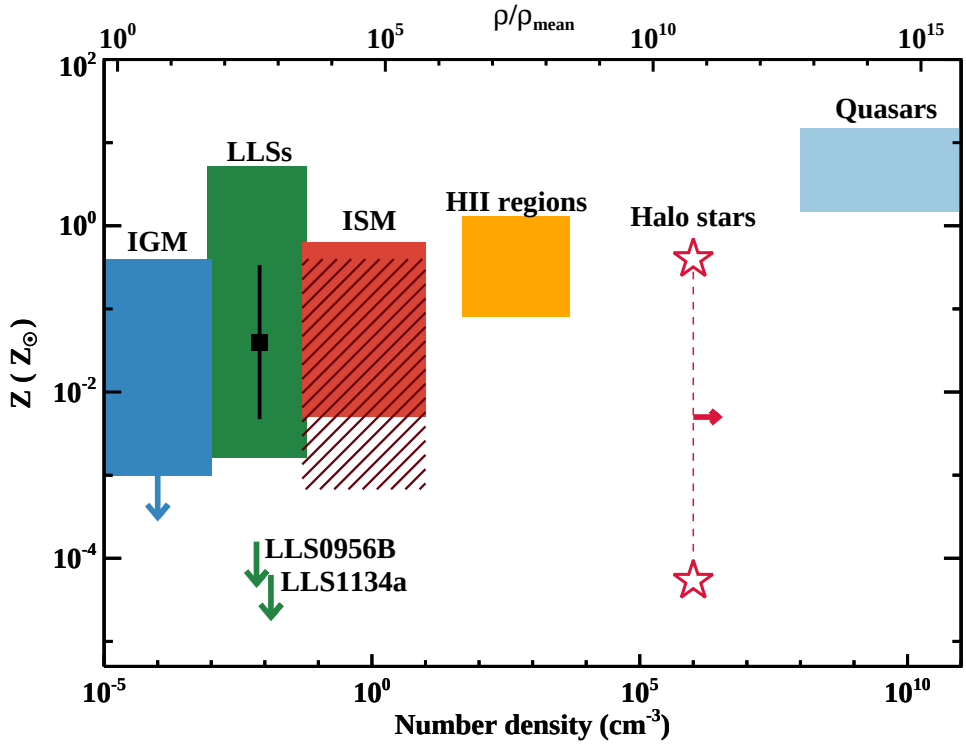

Figure 3.1: Peak to peak variation of the observed metallicity in $z \gtrsim 2$ cosmic structures at different densities. The blue, green and red rectangles show the spread in observed metallicities for diffuse gas in the universe; respectively, these are the IGM, LLSs, and galactic ISM. Orange rectangle: H II regions in galaxies. Light blue rectangle: quasar broad line regions. The black point with error bars marks the mean metallicity and the standard deviation for $z > 1.5$ LLSs. Galactic halo stars (stars connected with a dashed line) are represented at arbitrary density for visualization purposes. The top axis translates the number density in the overdensity above the mean baryon cosmic density at $z = 3.5$. The upper limits on the metallicity for LLS1134a and LLS0956B are shown with green arrows, assuming $\log U = -3$. Higher ionization parameters would shift these limits to lower densities and lower metallicity. See the SOM text 5 for additional details on the observations presented in this figure.

ital Sky Survey had shown significant absorption at wavelengths $\lambda < 4000$Å characteristic of substantial optical depth $\tau_{\rm LL}$ at the H I Lyman limit (at wavelength $\lambda < 912$Å), typical of the Lyman limit systems (LLSs).

Our Keck/HIRES spectrum of J1134+5742 reveals a sharp break in the flux at ~ 4000 Å, indicating the presence of a LLS at $z \sim 3.4$ with $\tau_{\rm LL} > 2$ (hereafter named LLS1134; Figure 3.2), confirming the lower resolution data. A search for absorption lines at a redshift consistent with the Lyman break further reveals the presence of the H I Lyman series through to Lyman–22 (H I 913.5), corresponding to $\lambda \sim 4030$Å. In the high resolution spectrum, two distinct absorbers can be identified within LLS1134, the main system (hereafter LLS1134a) at $z = 3.410883$ and a weaker component (named LLS1134b) at $z = 3.41167$, separated by $\delta v \sim 54$ km s^{-1}. In contrast, two flux decrements in the normalized HIRES spectrum of Q0956 + 122 (Figure 3.3) are visible at $\sim$ 3860 Å and $\sim$ 3750 Å, revealing the presence of two LLSs with $\tau_{\rm LL} \sim 1$ at $z \sim 3.22$ (LLS0956A) and $z \sim 3.10$ (LLS0956B).

A closer inspection of these spectra also reveals no detectable metal-line absorption at the velocity of the strongest H I component for neither LLS1134a nor LLS0956B. Such a complete absence of heavy element absorption has not been previously reported for data with comparable sensitivity. For LLS1134a, besides the H I Lyman series, the only other detected transitions are D I Lyα and Lyβ, offset by the appropriate $\delta v = -82$ km s^{-1} from the hydrogen absorption.

Table 3.1: Summary of the physical properties for LLS1134a and LLS0956B. For each system, we present: redshift, the hydrogen column density ($N_{\rm HI}$), the deuterium abundance (D/H), the Doppler parameter ($b_{\rm HI}$), the temperature, the metallicity, the hydrogen neutral fraction ($x_{\rm HI}$), the total hydrogen volume density ($n_{\rm H}$), and the ionization parameter (U)

	LLS1134a	LLS0956B
Redshift	3.410883 ± 0.000004	3.096221 ± 0.000009
$\log N_{\rm HI}$	17.95 ± 0.05	17.18 ± 0.04
$\log$ D/H	-4.69 ± 0.13^a	-

Continued on Next Page...

Table 3.1 – Continued

	LLS1134a	LLS0956B
$b_{\rm HI}$ (km s^{-1})	15.4 ± 0.3	20.2 ± 0.8
Temperature (K)	$< (1.43 \pm 0.05) \times 10^4$	$< (2.48 \pm 0.19) \times 10^4$
Metallicity ($Z_\odot$)	$< 10^{-4.2}$	$< 10^{-3.8}$
$\log x_{\rm HI}$	< -2.10	< -2.40
$\log n_{\rm H}$	< -1.86	< -1.98
$\log U^b$	> -3	> -3

To characterize the physical properties of these LLSs (summarized in Table 3.1), we modeled the hydrogen and deuterium absorptions using a χ^2 minimization algorithm (Table 3.3; Figure 3.7 and 3.8, see SOM text 2) and we derived upper limits on the column densities of various ionization states of heavy elements (Table 3.2; Figure 3.9, see SOM text 3). To translate these limits into constraints on the gas metallicity, it is necessary to make assumptions on the ionization state of the gas (see SOM text 4).

Table 3.2: 2σ upper limits on the column density of metal ions and rest-frame equivalent widths. Rest frame wavelengths and oscillator strengths (f) from Morton (2003) are also listed.

			LLS1134a		LLS0956B	
Ion	λ (Å)	$\log f$	$\log N$ (cm^{-2})	W_0 [a] (mÅ)	$\log N$ (cm^{-2})	W_0 [a] (mÅ)
C II	1334.5323	-0.8935	< 12.45	< 5.7	< 12.26	< 3.7
C III	977.0200	-0.1180	< 12.20	< 10.2	-	-
C IV	1548.1950	-0.7194	< 12.09	< 5.0	< 12.30	< 8.1
O I	1302.1685	-1.3110	< 12.76	< 4.2	< 12.50	< 2.3
Si II	1260.4221	0.0030	< 11.30	< 2.8	-	-
Si II	1526.7066	-0.8962	< 12.27	< 4.9	< 12.45	< 7.4
Si III	1206.5000	0.2201	< 11.40	< 5.4	< 11.21	< 3.5
Si IV	1393.7550	-0.2774	< 11.69	< 4.4	< 11.83	< 6.1
Fe II	1608.4511	-1.2366	< 12.59	< 5.2	< 12.72	< 7.0

[a] Including LLS1134a, the current best estimate for the primordial deuterium abundance becomes $\overline{\log({\rm D/H})} = -4.556 \pm 0.034$.

[b] The listed values are physically motivated, but not directly measured. Note that the metallicity, $x_{\rm HI}$, and $n_{\rm H}$ depend on the assumed value.

[a] The rest-frame equivalent width (W_0) is computed for the linear portion of the curve of growth.

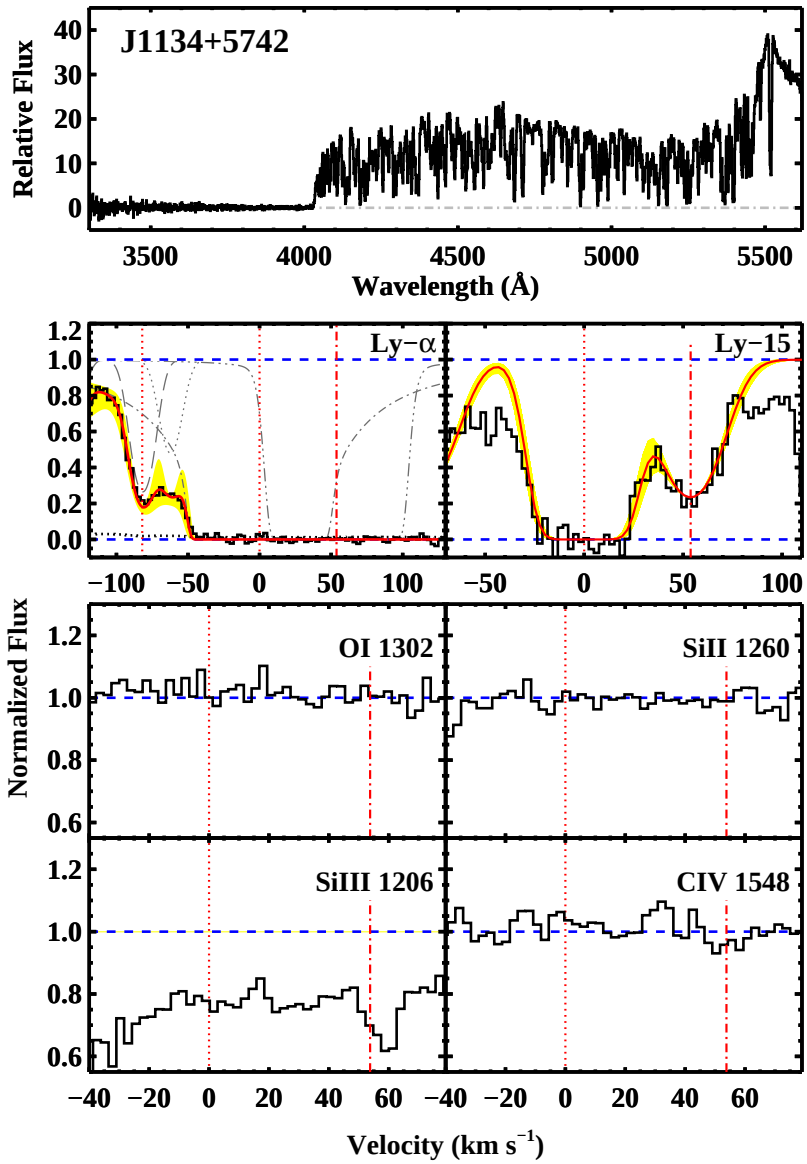

Figure 3.2: Top panel: Keck/LRIS spectrum of the QSO J1134+5742. A LLS at $z \sim 3.4$ is clearly visible from the break at $\sim$ 4000 Å. Middle panels: H I Lyman series transitions in the LLS1134a. Superimposed to the data are the best-fit model (red line) and the 2σ errors (yellow shaded regions). Individual components included in the model are marked with thin gray lines and the position of the hydrogen and deuterium are indicated by vertical dotted lines. The second hydrogen component (LLS1134b) at +54 km s^{-1} is marked with a dash-dotted line. Bottom panels: selected strong metal-line transitions in LLS1134a (dotted lines) and LLS1134b (dash-dotted lines). Unrelated absorption from the IGM contaminates the Si III transition (see Figure 3.9)

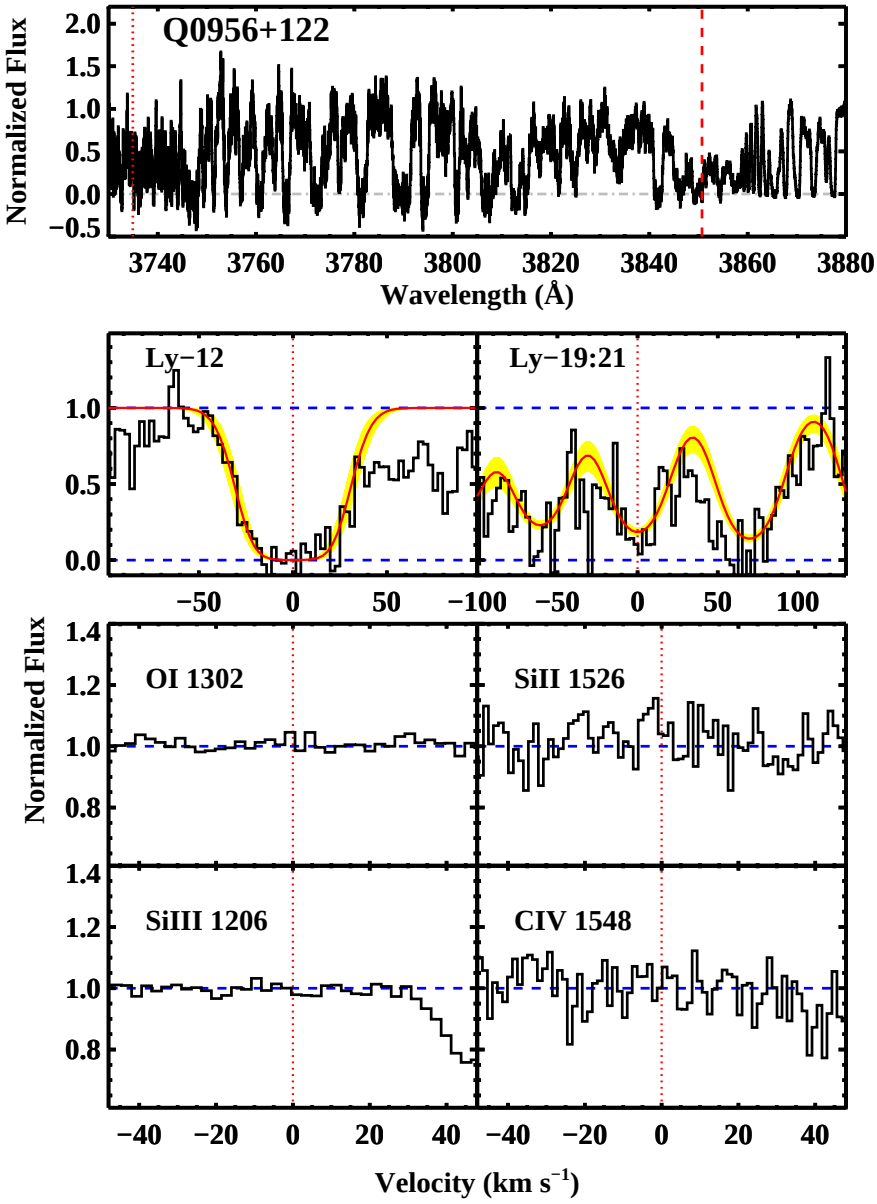

Figure 3.3: Top panel: Keck/HIRES spectrum of the QSO Q0956 + 122. The flux decrements at $\sim$ 3860 Å and $\sim$ 3750 Å reveal two partial LLSs at $z \sim 3.22$ (LLS0956A) and $z \sim 3.10$ (LLS0956B). The corresponding Lyman limits are marked with vertical lines. Middle panels: hydrogen Lyman series transitions for LLS0956B shown relative to $z = 3.096221$. Superimposed to the data, the best-fit model (red lines) and 2σ uncertainties (yellow shaded regions). Bottom panels: selected strong metal-line transitions for LLS0956B

Every LLS with $N_{\rm HI} \lesssim 10^{19}\,{\rm cm}^{-2}$ analyzed to date has exhibited absorption characteristic of a predominantly ionized gas (e.g. Prochter et al. 2010, Prochaska & Burles 1999). Standard interpretation is that the medium has been photoionized by an external radiation field, presumably a combination of the extragalactic UV background (EUVB) generated by the cosmological population of quasars and galaxies together with emission from local sources (e.g. a nearby galaxy). This conclusion is based on comparison of the observed ionic column densities of heavy elements with simple photoionization models. Every previous LLS has shown substantial absorption from doubly and triply ionized species, e.g. Si^{++} and C^{+3}, which trace predominantly ionized gas. Parametrizing the ionization state in terms of the ionization parameter $U \equiv \Phi/(c\,n_{\rm H})$, with Φ the flux of ionizing photons and $n_{\rm H}$ the gas volume density, all previously analyzed LLSs have exhibited $U \geq 10^{-3}$ corresponding to $n_{\rm H} \leq 10^{-2}\,{\rm cm}^{-3}$ at $z \sim 3$ (Figure 3.10).

Table 3.3: Summary of the best-fit model for the LLS1134 hydrogen and deuterium absorption. For each ion we list: the redshift of the absorption, the velocity offset relative to the main hydrogen component, the Doppler parameter and the column density.

Ion	$z_{\rm abs}$	δv (km s^{-1})	Doppler parameter (km s^{-1})	Column density ($\log N$)
D I	3.410883 ± 0.000004	0	10.2 ± 0.8	13.26 ± 0.04
H I	3.40997 ± 0.00002	-62	9.9 ± 1.9	12.99 ± 0.07
H I	2.7209 ± 0.0001	–	36.9 ± 9.2	13.05 ± 0.12
H I	3.410883 ± 0.000004	0	15.4 ± 0.3	17.94 ± 0.05
H I	3.41167 ± 0.00001	54	18.0 ± 0.9	16.71 ± 0.02

Adopting this U value as a limit to the ionization state and current estimates for the spectral shape of the EUVB (Haardt & Madau 2011), we infer metallicities of $Z < 10^{-4.2} Z_{\odot}$

and $Z < 10^{-3.8} Z_{\odot}$ for LLS1134a and LLS0956B (see Figure 3.11). These upper limits are 100 to 1000 times lower than typical measurements of LLSs and over an order of magnitude lower than any previous observed metallicity at $z > 2$ (Figure 3.1). These limits are only comparable to the abundances detected in the most metal poor star (Caffau et al. 2011) and are suggestive of a primordial composition. Super-solar metallicity is commonly found in the surroundings of quasars as well as in a few LLSs. Metallicities between solar and $\sim 1/10$ solar are typical of galaxies, and sub-solar metal enrichment down to $10^{-3} Z_{\odot}$ is characteristic of the ISM and IGM at early epochs. Remarkably, the most iron poor stars in the Galactic halo, thought to be the repository of the first generation of metals, have total heavy element abundance comparable to these limits, but generally above $\sim 1/1000$ solar. Our analysis uncovers regions of the universe at $z < 6$ with essentially primordial enrichment, whose traces can be found within the oldest stellar populations in the present universe (Caffau et al. 2011).

The detection of deuterium in the metal free LLS1134a provides a direct confirmation of the BBN. From the analysis of the D and H absorption lines, we derive $\log(\mathrm{D/H}) = -4.69 \pm 0.13$ (see SOM text 2). The observed value is consistent at 1σ with the theoretical predicted value $\log(\mathrm{D/H}) = -4.592$ (Steigman 2007), assuming $\Omega_{b,0}$ from WMAP7. This measurement is in agreement with previous determinations in quasar absorption line systems at $z > 2$ (Table 3.4; Figure 3.4). Because the lack of metals in LLS1134a confirms its pristine composition, this agreement strengthens the hypothesis that at low metallicities ($10^{-2} Z_{\odot}$ or less) the observed deuterium abundances are representative of the primordial value (Kirkman et al. 2003) and astration cannot be responsible for the lingering scatter in the observed D/H. Deuterium abundances from quasar absorption line systems are therefore solid anchor points for models of galactic chemical evolution (e.g. Romano et al. 2006). Combining D/H in LLS1134a with values from the known D-bearing systems, we obtain a logarithmic weighted mean $\overline{\log(\mathrm{D/H})} = -4.556 \pm 0.034$ that translates into $\Omega_{b,0} h^2(\mathrm{BBN}) = 0.0213 \pm 0.0012$, after accounting for both random and sys-

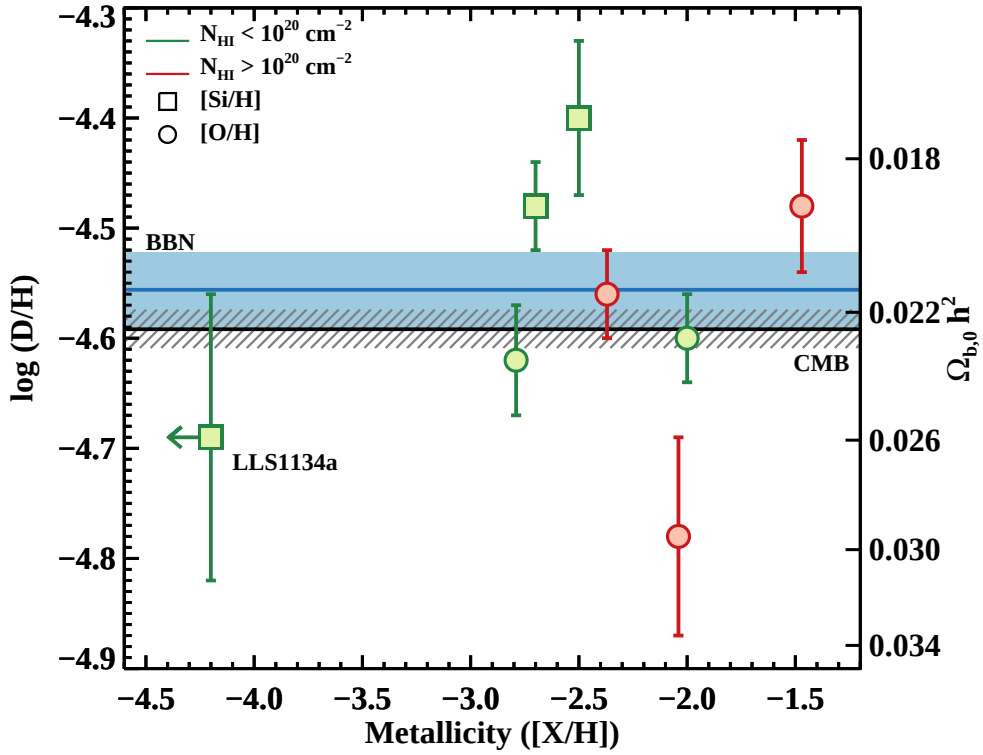

Figure 3.4: Deuterium abundances as a function of metallicity ($[\mathrm{X/H}] \equiv \log(\mathrm{X/H}) - \log(\mathrm{X/H})_\odot$) for the $z > 2$ absorption line systems. Green symbols are for LLSs, while red symbols are for higher column density absorbers ($N_{\mathrm{HI}} > 10^{20}$ cm^{-2}). Metallicities obtained with silicon are indicated by squares, while those obtained with oxygen are indicated with circles. The right-hand axis translates the deuterium abundance to the cosmic baryon density $\Omega_{b,0}h^2$. The inferred $\Omega_{b,0}h^2$(BBN) is shown with a solid blue line, together with the 1σ errors (light blue shaded area). The CMB value and 1σ errors from WMAP7 are instead shown with a solid black line and a gray dashed area.

tematic errors (Pettini et al. 2008b). Consistent with previous studies, we do not include in the weighted mean the error on the assumed level of the quasar continuum light.

Table 3.4: Deuterium abundances in $z > 2$ absorption line systems[e].

Quasar	$z_{\rm abs}$	log D/H	log $N_{\rm HI}$	[X/H]
HS0105 + 1619	2.53600	-4.60 ± 0.04	19.42 ± 0.01	-2.00^a
Q0913 + 072	2.61843	-4.56 ± 0.04	20.34 ± 0.04	-2.37^a
Q1009 + 299	2.50357	-4.40 ± 0.07	17.39 ± 0.06	-2.5^b
J1134+5742	3.41088	-4.69 ± 0.13	17.95 ± 0.05	$< -4.2^{b,c}$
Q1243 + 3047	2.52566	-4.62 ± 0.05	19.73 ± 0.04	-2.79^a
SDSSJ1558 − 0031	2.70262	-4.48 ± 0.06	20.67 ± 0.05	-1.47^a
Q1937 − 1009	3.57220	-4.48 ± 0.04	17.86 ± 0.02	$-2.7, -1.9^{b,d}$
Q2206 − 199	2.07624	-4.78 ± 0.09	20.43 ± 0.04	-2.04^a

The absence of metals in LLS1134a and LLS0956B is outstanding also in the framework of theories for the metal enrichment of cosmic structures (Figure 3.5). Numerical simulations suggest that LLSs typically arise in galaxies (e.g. Kohler & Gnedin 2007), in dense gas above ~ 100 times the mean cosmic density $\rho_{\rm mean}$. At the same time, in models of the IGM enrichment, metals are ejected to hundreds of kpc from star forming regions, resulting in substantial pollution of the nearby gas. As a consequence, the metallicity predicted for stars or for the ISM at $z < 4$ (Oppenheimer et al. 2012) ranges between $0.1 - 1 Z_{\odot}$, three orders of magnitude higher than the limits inferred for these two LLSs. Similarly, the metallicity predicted for the hot halo of galaxies and for the surrounding IGM (Oppenheimer et al. 2012, Hernquist & Springel 2003) exceeds by a factor of 10 or more the limits for LLS1134a and LLS0956B. Contrary to any prediction and any previous observation, these two LLSs reside at significant overdensity ($\rho/\rho_{\rm mean} < 850$) but

[a] Metallicity obtained using oxygen.
[b] Metallicity obtained using silicon.
[c] This work.
[d] The authors report metallicities for two components.
[e] Data from O'Meara et al. (2006), Pettini et al. (2008b), Kirkman et al. (2003), O'Meara et al. (2001), Burles & Tytler (1998;?), Pettini & Bowen (2001), Pettini et al. (2008a).

in an unpolluted portion of the universe.

Metallicity below $10^{-4} Z_{\odot}$ appears even exceptional when compared to the level of pre-enrichment from PopIII stars that is predicted by models (Mackey et al. 2003, Yoshida et al. 2004, Wise et al. 2012) already 500 million years after the Big Bang ($z \sim 10$). Our limits place additional constraints on the widespread dispersal of metals from primordial stellar populations and the first generations of galaxies. Remarkably, the gas we detected could in principle fuel PopIII star-formation at $z \sim 3$, because its metallicity lies at or even below the minimum enrichment required for metal cooling to induce fragmentation in the collapsing material in the absence of dust (Bromm et al. 2001). Given a lower limit on the LLS physical size $\ell = N_{\rm HI}/(x_{\rm HI} n_{\rm H})$, with $x_{\rm HI}$ the neutral fraction and $n_{\rm H}$ the total hydrogen volume density, we can infer the total hydrogen mass in these clouds $M_{\rm H} = m_{\rm p} N_{\rm H} \ell^2$. We find $M_{\rm H} \gtrsim 6.2 \times 10^6$ M$_{\odot}$ for LLS1134a and $M_{\rm H} \gtrsim 4.2 \times 10^5$ M$_{\odot}$ for LLS0956B, comparable to the mass of the mini-halos where the first generation of stars formed (e.g. Wise et al. 2012). Therefore, if this gas were able to collapse further and shield from the ambient UV radiation, it would potentially give rise to PopIII stars two billion years after the transition between the first and second generation of stars (PopII) is thought to have occurred. Thus, pair-production supernovae associated with the death of these massive and metal-free stars may be found even at modest redshifts (Scannapieco et al. 2005).

The pristine composition of LLS1134a and LLS0956B can be reconciled with model predictions and previous observations if mixing of metals within the IGM is an inefficient and inhomogeneous process. A varying degree of metal enrichment is seen in multiple components of LLSs (Prochter et al. 2010, Prochaska & Burles 1999, D'Odorico & Petitjean 2001), implying that mixing does not operate effectively on small scales. The detection of ionized metals in LLS1134b, the weaker component at +54 km s^{-1} from LLS1134a, reinforces this point. Further, studies of metal systems in the low density and diffuse IGM suggest that at least some of the ionized heavy

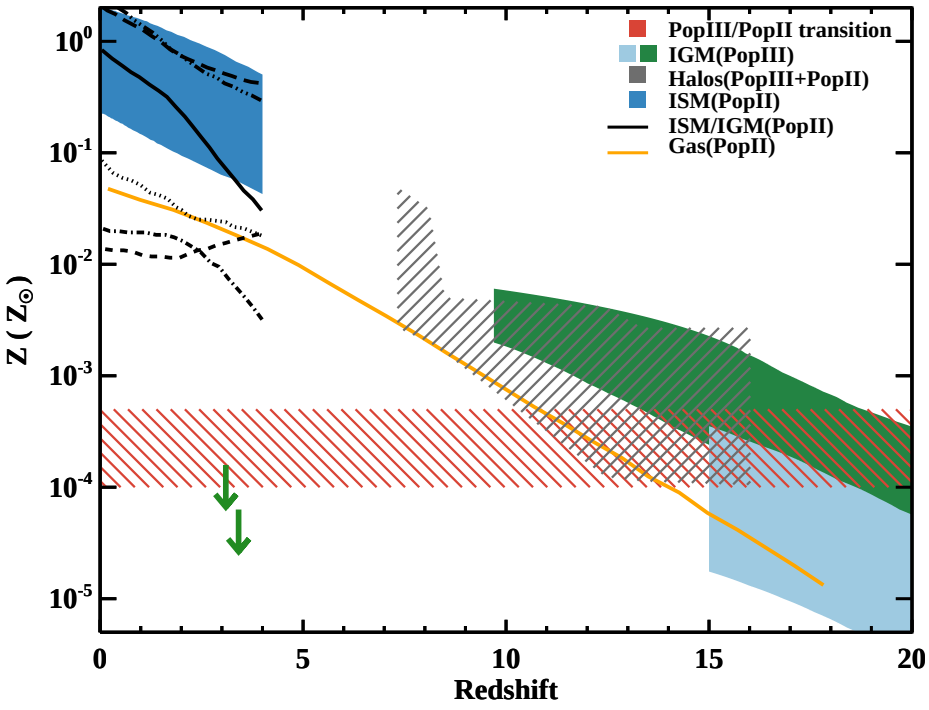

Figure 3.5: Overview of analytic models and simulations for the metal enrichment of the universe. Light blue and green shaded regions: IGM metallicity from PopIII stars with mixing between $1 - 0.05$ and different star formation histories (Mackey et al. 2003, Yoshida et al. 2004). Red dashed region: critical metallicity that marks the transition between PopIII and PopII stars (Bromm et al. 2001, Schneider et al. 2002). Orange line: analytic model for the gas metal content in the universe from PopII stars and galactic winds (Hernquist & Springel 2003). Gray shaded region: gas metallicity within halos from hydrodynamical simulations that include yields from both PopIII and PopII stars (Wise et al. 2012). Blue shaded region: analytic model for the ISM metallicity at different halo masses ($10^{11} - 10^{14} M_{\odot}$) and different wind models (Davé et al. 2012). Black lines: metallicity from hydrodynamical simulations with momentum driven winds (Oppenheimer et al. 2012) in condensed gas (solid line), hot halo (dotted line), warm-hot intergalactic medium (dashed line), diffuse gas (dash-dotted line), ISM (dash-triple-dotted line), and stars (long-dashed line). Upper limits on the metallicities of LLS1134a and LLS0956B are marked with green arrows.

elements (e.g. C IV) are in small and short-lived clumps (Schaye et al. 2007). A low volume filling factor for metals is also consistent with theories of metal ejection from supernovae, in which most of the heavy elements are initially confined in small bubbles (Ferrara et al. 2000) and only subsequently diffuse in the surrounding IGM. Plausibly, LLS1134a and LLS0956B originate in a filament of the cosmic web where primordial regions coexist with enriched pockets of gas. If the metal enrichment is highly inhomogeneous, these two LLSs could just be the tip of the iceberg of a much larger population of unpolluted absorbers that trace a large fraction of the dense IGM.

Beside the implications for the BBN and the metal distribution, the detection of metal free LLSs is tantalizing in the context of galaxy formation and evolution. Modern theory and simulations predict that most of the gas that sustains star formation is accreted in galaxies through dense and narrow streams, known as cold flows (Kereš et al. 2005, Dekel et al. 2009b). These gaseous filaments are highly ionized by both the EUVB and the radiation escaping from the central star forming regions. Cold flows should therefore appear as LLSs in the spectra of bright quasars (Faucher-Giguère & Kereš 2011, Fumagalli et al. 2011a, van de Voort et al. 2012). Further, in contrast to metal enriched gas that is outflowing from galaxies, this infalling material is expected to be metal poor (Fumagalli et al. 2011a). Although direct observational evidence of cold flows is still lacking, these streams are thought to be ubiquitous at high-redshift and primordial LLSs such as LLS1134a and LLS0956B are ideal candidates for this elusive mode of accretion.

3.3 Supplementary material

3.3.1 Observations and data reduction

Previous spectra of quasars SDSS J113418.96+574204.6 and Q0956 + 122 had shown substantial absorption at wavelengths $\lambda < 4000$Å characteristic of the continuum opacity of H I gas. LLS1134a and LLS0956B were identified as part of an ongoing program to explore the incidence and physical characteristics of LLSs (Prochter et al. 2010, O'Meara et al. 2007) from a sample of ~ 50 absorbers, many of which have apparently low metallicity ($< 1/100$ solar).

We first observed quasar J1134+5742 on UT January 5, 2006 with the HIRES spectrometer (Vogt et al. 1994) on the Keck I telescope on Mauna Kea. The data were obtained in two exposures totaling 6300 seconds, with the instrument configured with the red cross-disperser, the C5 decker, and the kv380 blocking filter to suppress second order flux. In this configuration, the instrument provides a resolution of ~ 8 km s^{-1} FWHM, and the echelle and cross-disperser angles were set to provide wavelength coverage 4017 Å $< \lambda <$ 8534 Å. These spectra have gaps in coverage at $\lambda \sim 5450$Å and 7050Å due to spacings between the CCDs in the detector mosaic. There are also a series of gaps in wavelength coverage beyond $\lambda > 6300$Å where the free spectral range of the spectrometer exceeds the detector width.

We observed Q0956 + 122 with HIRES once on UT January 3, 2006 for 7200 seconds, and again on UT April 7, 2006 for 1800 seconds. For the January data, HIRES was configured with the UV cross-disperser and the C5 decker, again providing ~ 8 km s^{-1} FWHM resolution. The wavelengths covered in this configuration were 3621 Å $< \lambda <$ 5533 Å, with gaps in coverage only related to the CCD mosaic. The April 2006 data had HIRES configured with the red cross-disperse, the C1 Decker, and the og530 blocking filter. The wavelengths 5465 Å $< \lambda <$ 9998 Å were covered with wavelength gaps due to the spacings of the CCD mosaic and the limited detector size.

For both quasars, the 2D images were reduced with the HIRESredux pipeline[1] which extracts and coadds the data. An interactive continuum fitting procedure within the pipeline assigns a continuum level to each coadded spectral order, and the continuum normalized spectra are then combined into a single 1D spectrum. For J1134+5742, the resultant 1D spectrum has a signal-to-noise ratio S/N $\sim$ 25 per 2.6 km s^{-1} pixel at 6300Å, the central wavelength. For Q0956 + 122, the 1D spectra have S/N $\sim$ 30 per 2.6 km s^{-1} pixel at 4500Å for the UV cross-disperser data, and $\sim$ 10 per 1.4 km s^{-1} pixel at 6000 Å for the red cross-disperser data.

We also observed J1134+5742 on UT 2011 July 1 with the LRIS spectrograph (Oke et al. 1995) on the Keck I telescope on Mauna Kea. We obtained a single exposure of 400 seconds, with the instrument configured with the 600/4000 grism, the D560 dichroic and the 600/7500 grating, tilted to ensure continuous spectral coverage between the blue and red arms. The 2D images were reduced with the LowRedux pipeline[2] which calibrates and extracts the data.

3.3.2 Hydrogen and deuterium analysis

We now describe the H I and D I analysis for each LLS.

3.3.2.1 The Lyman limit system at $z = 3.410883$ toward J1134+5742

Both the strong Lyman limit absorption visible in the Keck/LRIS spectrum (Figure 3.2), and the saturation of the high-order Lyman series lines in the Keck/HIRES data (Figure 3.7) imply a large H I column density, n_{HI}. Figure 3.6 shows a zoom-in of the Keck/LRIS spectrum for J1134+5742, focusing on the Lyman limit absorption. We estimate the flux just redward of the Lyman limit to be 13 ± 2 in our relative units. This estimate appropriately includes the attenuation of the quasar light by the Lyα forest. In the following, we assume a flat spectrum for the continuum and note that a more realistic spectrum would yield a slightly

[1] http://www.ucolick.org/~xavier/HIRedux/index.html
[2] http://www.ucolick.org/~xavier/LowRedux/index.html

higher estimate for the n_{HI} value. Just blueward of the Lyman limit ($\lambda = 3950 - 4000$Å), we set a conservative upper limit to the relative flux of 0.05. Combining these two measurements, we estimate a lower limit to the optical depth at the Lyman limit of $\tau_{LL} > \ln(9./0.05) > 5$ for a total $N_{HI} > 10^{17.9}\,cm^{-2}$. The model for this n_{HI} limit, shown as the green line in Figure 3.6, is marginally acceptable at $\lambda \sim 4000$Å, but predicts a recovery of the flux at $\lambda < 3800$Å that appears inconsistent with the observations. Metal-lines and strong Lyα lines in the HIRES spectrum, however, reveal the presence of other absorbers along the sightline to J1134+5742 at redshifts $z = 3.0283$, $z = 3.0753$, $z = 3.1047$, $z = 3.2396$, $z = 3.2616$, and $z = 3.3223$. These systems could contribute to the opacity at $\lambda < 3942$Å. If they do not, then LLS1134a must have a higher n_{HI} value.

Figure 3.6 shows the data binned with a weighted mean in windows of 50Å between $3400 - 4000$Å. Overplotted on these binned evaluations (whose errors only reflect statistical uncertainty) is a model curve for $N_{HI} = 10^{18.05}\,cm^{-2}$ (blue line). This n_{HI} value reproduces the observed absorption to ~ 3400Å. Larger n_{HI} values are not ruled out by the analysis, especially when one allows for systematic error (e.g. sky subtraction). In summary, we report a lower limit to the total n_{HI} value of $10^{17.9}\,cm^{-3}$ and note the data require yet higher values if LLS1134a dominates the opacity at $\lambda < 3700$Å.

To further refine the n_{HI} measurement and its redshift distribution, we have performed detailed analysis of the Lyman series lines. This analysis, however, is challenged by the fact that gas with $N_{HI} \sim 10^{18}\,cm^{-2}$ has Lyman series lines that lie on the saturated portion of the curve-of-growth (e.g. Prochter et al. 2010). As such, we proceed cautiously as this may impose large uncertainties in the model. We may first set a strict upper limit to the total n_{HI} of LLS1134a from the absence of substantial damping wings in the Lyα profile (Figure 3.2). Because the flux would be over-absorbed for all models with n_{HI} exceeding $10^{18.7}\,cm^{-2}$, this is the largest n_{HI} value allowed for LLS1134a. Second, the complete absorption at $\lambda \sim 4030$Å by the Lyman

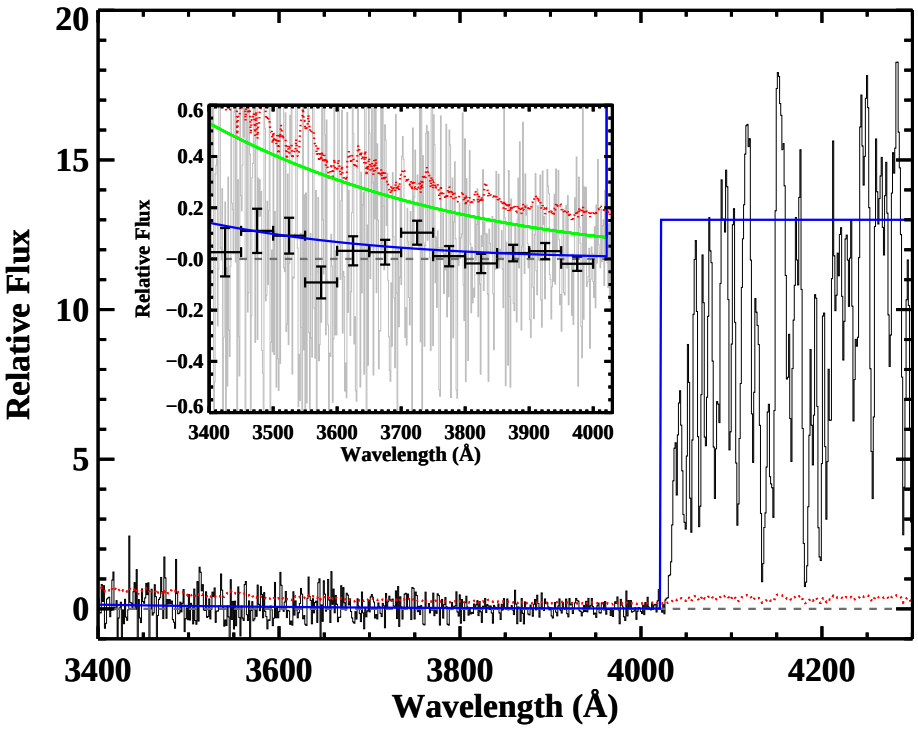

Figure 3.6: LRIS spectrum of J1134+5742 at the Lyman limit. Superimposed to the data, we show the flux decrement for a $N_{\rm HI} = 10^{18.05}\,{\rm cm}^{-2}$ LLS (blue line) together with the error on the flux (red dotted line). In the inset, we show the data at full resolution (gray histogram) and in bins of 50 Å (crosses). A model of the Lyman limit opacity for $\log N_{\rm HI} > 17.90$ (green line) is also shown.

series requires $N_{\rm HI} > 10^{17.6}\,{\rm cm}^{-2}$, consistent with the Lyman limit analysis performed above.

Lastly, we can model the H I absorption using the VPFIT V9.5 package that allows one to fit multiple Voigt profiles to the spectrum. Our general approach is to include as few components as possible (each defined by a redshift z, column density $\rm n_{HI}$, and Doppler parameter $b_{\rm HI}$) to match the Lyman series absorption. The data are best described by two principal H I components (LLS1134a and LLS1134b), as evident from a visual inspection of the higher order Lyman series lines (Figure 3.2 and 3.7). Since LLS1134b is unsaturated beyond Ly-14, we first measure the line-parameters of this subsystem using the Lyman series lines between Ly-10 and Ly-16. We find $z = 3.41167 \pm 0.00001$, $b_{\rm HI} = 18.0 \pm 0.9$ km s^{-1} and $\log N_{\rm HI} = 16.71 \pm 0.02$ indicating that this component has a nearly negligible contribution to the observed Lyman limit. Adding an additional uncertainty of 0.02 dex related to the $1-3\%$ error in the continuum placement, our best fit value for the column density of LLS1134b becomes $\log N_{\rm HI} = 16.71 \pm 0.03$. Then, we model LLS1134a and the deuterium absorption, using both high order Lyman series lines (Ly-7 through Ly-16) as well as Ly-α and Ly-β. For these transitions, we must include absorption from gas unrelated to LLS1134 which we ascribe to coincident Lyα absorption from lower redshift absorbers. A summary of the model parameters is given in Table 3.3.

The strongest H I component has $z = 3.410883 \pm 0.000004$, $\log N_{\rm HI} = 17.94 \pm 0.05$ and an H I Doppler parameter $b_{\rm HI} = 15.4 \pm 0.3$ km s^{-1}. The latter is consistent with a predominantly ionized gas at a temperature $T \leq 1.4 \times 10^4$ K. The line-parameters for LLS1134a are well constrained by the full Lyman series and the resultant $\rm n_{HI}$ value is consistent with our Lyman limit analysis of the Keck/LRIS spectrum. Despite the line-saturation, there is substantial constraint on the $b_{\rm HI}$ value, and thereby the $\rm n_{HI}$ value, from the line-profile shapes of the Lyman series lines. This conclusion hinges, however, on the assumption that the H I absorption at $z \sim 3.41088$ is dominated by a single component. As we will demonstrate in the next paragraph, this assumption is well supported by the analysis of neighboring D I absorption.

From the above discussion, we conclude $N_{\rm HI} = 10^{17.95\pm0.05}\,{\rm cm}^{-2}$ for LLS1134a. The 0.02 dex error on the continuum placement is in this case negligible compared to the statistical uncertainty. In regards to the metallicity of this gas, we further emphasize that $\log N_{\rm HI} \geq 17.95$ is preferred by the analysis of the Lyman continuum opacity, despite the symmetric statistical error. If anything, metallicity limits derived from this $n_{\rm HI}$ value may be considered conservative.

In addition to the H I Lyman series, LLS1134a exhibits substantial absorption for the D I Lyα and Lyβ transitions at the expected -82 km s^{-1} offset from the H I lines. Due to blending with absorption lines from lower redshift gas, deuterium is not detected in the remaining Lyman series lines. Although both of the D I transitions are partially blended with (presumed) H I absorption, a simultaneous fit to the hydrogen and deuterium absorption lines (Figure 3.7) constrains $\log N_{\rm DI} = 13.26 \pm 0.04$ and $b_{\rm DI} = 10.2 \pm 0.8$. The latter parameter is fully consistent with the $b_{\rm HI}$ value derived from the hydrogen absorption. The deuterium estimate is more sensitive to the continuum placement error than hydrogen, particularly at the position of Ly-β. Therefore, we include an additional uncertainty of ~ 0.08 dex, computed by repeating multiple times this analysis after having varied the continuum level by $\pm 1 - 3\%$. When we express the Doppler parameter as a function of thermal broadening $b_{\rm ther}$ and turbulent component $b_{\rm turb}$, the expected ratio of the hydrogen and deuterium b parameters is $b_{\rm DI}/b_{\rm HI} = \sqrt{1 - 0.5\phi}$ with $\phi = b^2_{\rm HI,ther}/(b^2_{\rm HI,ther} + b^2_{\rm turb})$ and $0 < \phi < 1$. The measured $b_{\rm DI}/b_{\rm HI} = 0.66$ is within 10% of the expected value for turbulent broadening ($\phi \sim 1$). The best-fit D/H ratio for LLS1134a is $\log({\rm D/H}) = -4.69 \pm 0.06$, in agreement with previous determinations (Table 3.4). This consistency provides additional confidence on the model for the hydrogen absorption. Due to the partial blending of the D I lines with intervening hydrogen, the deuterium column density is affected by the additional uncertainty on the line parameters for the hydrogen absorption. Including this error, our best estimate becomes $\log({\rm D/H}) = -4.69 \pm 0.09$. Finally, including the uncertainty on the continuum determination, we find $\log({\rm D/H}) = -4.69 \pm 0.13$.

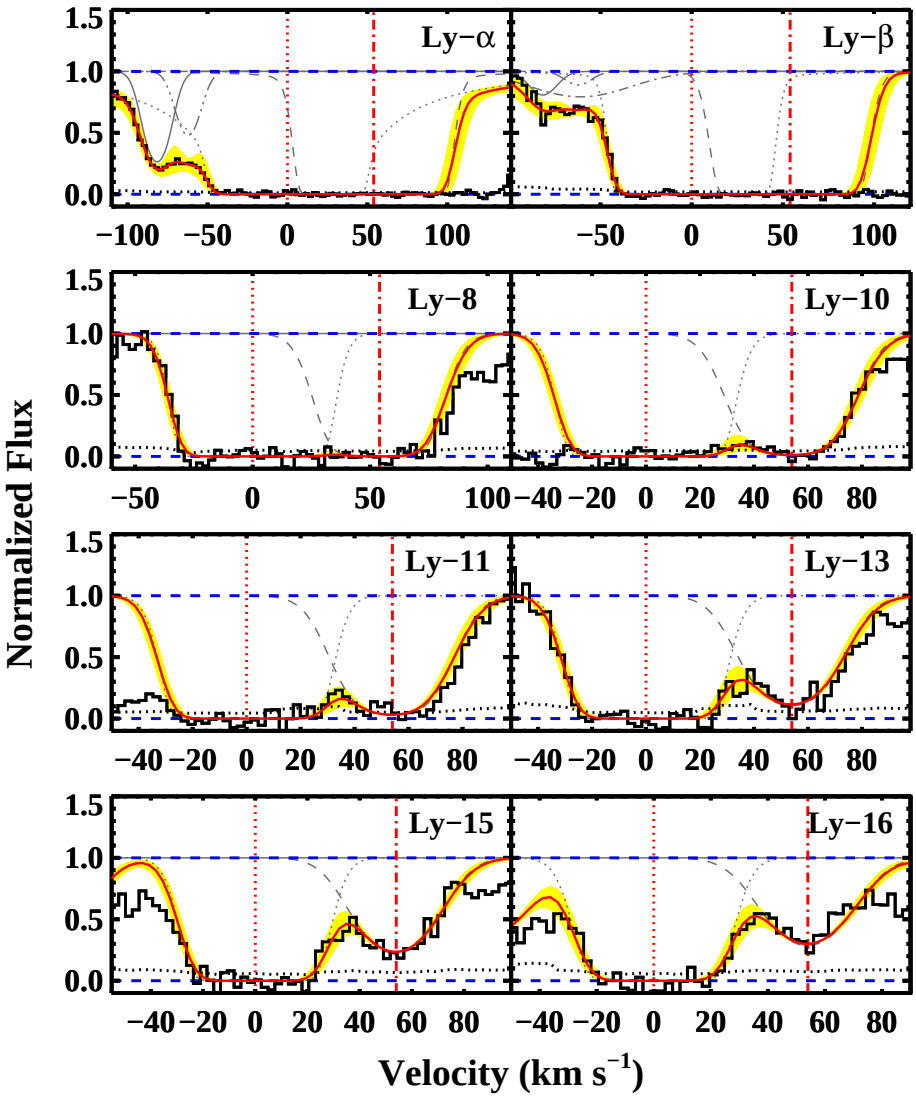

Figure 3.7: Gallery of the Lyman series transitions (black histograms) and best fit models (red lines) for LLS1134. The yellow shaded regions represent the 2σ errors on the line parameters. LLS1134a and LLS1134b are marked by vertical lines, while the dotted black lines indicate the 1σ error on the flux. Individual components included in the model are shown with gray lines.

3.3.2.2 The Lyman limit system at $z = 3.096221$ toward Q$0956 + 122$

A partial LLS (LLS0956A) was previously identified in low-resolution spectra of Q0956+ 122 and associated with corresponding metal-line absorption at $z = 3.2228$ (Steidel 1990). We confirm this identification (Figure 3.3). An LLS at lower redshift was also reported bySteidel (1990), who proposed it was associated with a metal-line system at $z = 3.1142$. Our analysis of the HIRES spectra, however, only reveals substantial high-order Lyman series absorption for a system at $z \sim 3.096$ and we adopt this as the redshift for LLS0956B.

Non-zero flux is detected blueward of the Lyman limit for each system, implying $N_{\rm HI} < 10^{17.8}\,{\rm cm}^{-2}$ for both LLS0956A and LLS0956B. At this optical depth ($\tau_{\rm LL} \sim 1$), the total $n_{\rm HI}$ value is well established by the Lyman limit opacity and the analysis of unsaturated Lyman series lines.

The hydrogen Lyman series of LLS0956A is detected through to Lyman-21 in our HIRES data, together with C IV and Si IV metal lines. By modeling the unsaturated hydrogen transitions from Lyman-16 to Lyman-21, we derive the redshift $z = 3.223194 \pm 0.000002$, the column density $\log N_{\rm HI} = 17.37 \pm 0.01$, and the Doppler parameter $b_{\rm HI} = 20.4 \pm 0.2$. Similarly, for LLS0956B, the hydrogen Lyman series is clearly visible through to Lyman-15 and, at lower signal-to-noise, through to Lyman-21. No other metal lines are detected at the hydrogen position, while deuterium is blended in the strongest transitions. Combining both high-order saturated and non-saturated H I absorption lines, we model this LLS with a single component at $z = 3.096221 \pm 0.000009$, with $b_{\rm HI} = 20.2 \pm 0.8$ km s^{-1} and $\log N_{\rm HI} = 17.18 \pm 0.03$ (Figure 3.3 and 3.8). After accounting for an additional 0.02 dex of uncertainty on the continuum placement, the best fit column density for LLS0956B becomes $\log N_{\rm HI} = 17.18 \pm 0.04$. The Doppler parameter implies a temperature $T \leq 2.5 \times 10^4$ K, typical for photoionized gas.

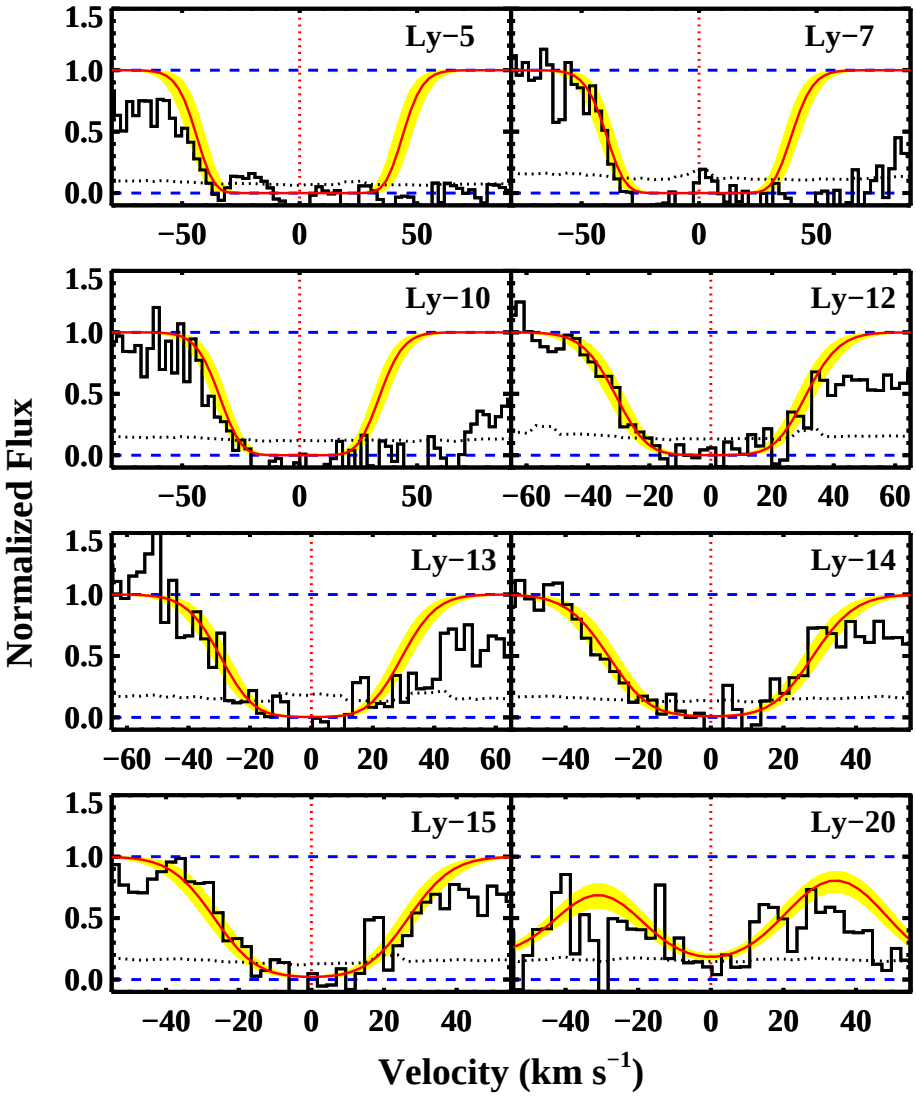

Figure 3.8: Same as Figure 3.7, but for LLS0956B.

3.3.3 Metal line analysis

We now discuss upper limits to the column densities of atoms and ions for heavy elements in these metal free LLSs.

3.3.3.1 LLS1134

LLS1134a is conspicuously free of metal-line absorption, as shown in Figure 3.2 where we mark for selected ionic transitions the position corresponding to this LLS and to the weaker hydrogen component ($z = 3.410883$ and $z = 3.41167$ respectively). For both LLS1134a and LLS1134b, we can place upper limits to the ionic column densities using the apparent optical depth method (Savage & Sembach 1991), with the key limiting factors being the S/N of the spectrum and contamination from the Lyα forest (i.e. Lyα lines at unrelated wavelengths). For this analysis, we choose a velocity window of ± 15 km s^{-1} that is wide enough to encompass the expected width for metal lines given the hydrogen and deuterium absorption properties. A larger velocity window of ± 20 km s^{-1} would yield slightly higher (~ 0.06 dex) limits. A summary of the metal column densities is presented in Table 3.2.

For most of the listed ionic transitions, we place robust upper limits based solely on the variance and the uncertainty associated with a $1-3\%$ (S/N dependent) error in continuum placement. Transitions blueward of 5450Å, however, lie within the Lyα forest which complicates the estimates. Of particular interest to the metallicity of LLS1134a are the strong transitions of Si III 1206 and C III 977. Figure 3.9 reveals that there is substantial absorption at the predicted wavelength for each transition. This absorption is broad, however, and has peak optical depth that is substantially offset from the predicted line-center. It is certain, therefore, that these transitions are blended with coincident absorption by the IGM. In this case, an upper limit based on the apparent optical depth would overestimate the column densities.

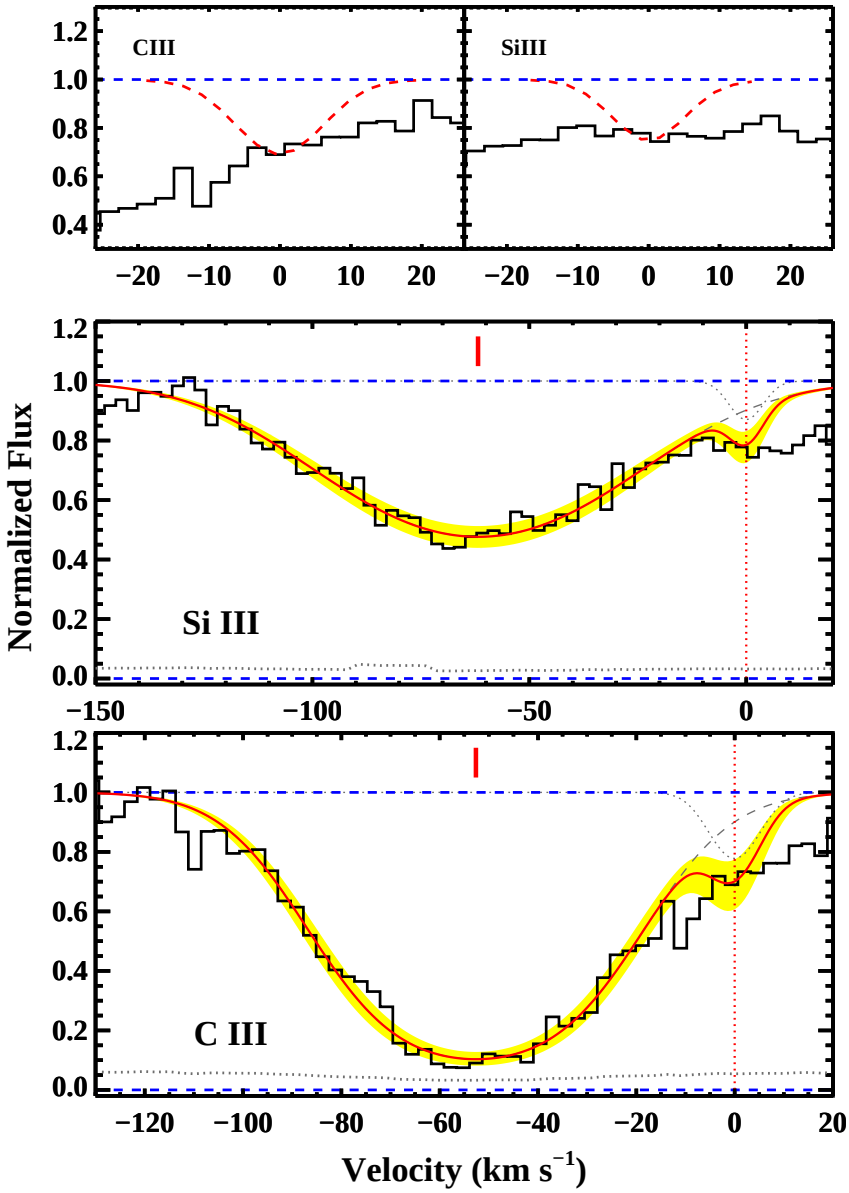

Figure 3.9: Top panels: C III 977 and Si III 1206 transitions for LLS1134a. Together with the HIRES data (black histograms), we display a model for the strongest lines that could be hidden by the IGM absorption (red dashed lines). In the bottom panels, we present a single component model (gray dashed line) for the IGM absorption that contributes to the opacity at the C III and Si III frequencies. Models for these transitions with the adopted column density limits are also shown (gray dotted line). In red, we display the combined two-component models, together with the corresponding 2σ errors on the line parameters. In all cases, we assume a Doppler parameter of 6.9 km s^{-1} for carbon and 5.3 km s^{-1} for silicon, as inferred from the hydrogen and deuterium lines.

Better limits to the Si III and C III column densities can be derived using models for these absorption lines. We fix the Doppler parameters for the metal lines using the hydrogen and deuterium b values. From the H I and D I lines, we infer a temperature $(1.6 \pm 0.2) \times 10^4$ K and an upper limit on the turbulent velocity $b_{\rm turb} < 2.2$ km s^{-1}. Carbon and silicon absorption lines that arise in gas with these properties would be characterized by $b_{\rm C} = 6.9$ km s^{-1} and $b_{\rm Si} = 5.3$ km s^{-1}. These Doppler parameters are within the range commonly observed for metal-line transitions (e.g. Lidz et al. 2010). A first limit to the column densities of C III and Si III is set by demanding that the peak optical depth of each metal line does not exceed the observed absorption. This approach, that basically ignores the fact that the opacity is dominated by coincident IGM lines, yields upper limits of $\log N_{\rm SiIII} < 11.85$ and $\log N_{\rm CIII} < 12.45$. Lines of column density higher than these would exceed the observed absorption at the corresponding transitions, as shown in the top panels of Figure 3.9.

A substantial fraction of the opacity at the Si III and C III frequencies must however arise from the two broad Lyα forest lines blueward to these metal transitions. We set therefore limits to $\log N_{\rm SiIII}$ and $\log N_{\rm CIII}$ using a two component model that includes both the metal lines and the Lyα forest. Following this procedure, we find $\log N_{\rm SiIII} < 11.40$ and $\log N_{\rm CIII} < 12.20$. Since additional Ly$\alpha$ lines at both negative and positive velocities are required to account for all the observed absorption, the choice of this minimal scenario (only two components) yields a quite conservative limit to the metal column densities.

Inspecting the strongest transitions of LLS1134b, the second main hydrogen component at $z = 3.41167$, we see that Si III and C III are detected. These transitions are commonly found in highly ionized LLSs. Again using the apparent optical depth method, we find for the column densities of these ions $\log N_{\rm CIII} = 13.03 \pm 0.02$ and $\log N_{\rm SiIII} = 12.15 \pm 0.02$. While a strong line is clearly visible at the position of S III for this component, the implied column density would greatly exceed that for Si III and C III indicating that the absorption is dominated by

contamination from the Lyα forest.

3.3.3.2 LLS0956B

Similarly to LLS1134a, LLS0956B is free of metal line absorption (Figure 3.3). Using the apparent optical depth method, we determine upper limits to the ionic column densities in the velocity interval ± 15 km s^{-1}, consistent with the expected metal line width. The metal column densities for the strongest transitions, including the variance and uncertainty related to continuum placement, are listed in Table 3.2.

C III λ977 is blended in the Lyα forest. Supported by the lack of any silicon absorption (Si II, Si III and Si IV) and the absence of C II and C IV, we conclude that carbon is not present in this system to limits comparable to the ones inferred from silicon. Therefore, we do not include C III λ977 in our estimates of the metallicity, but we rely on the Si III upper limit which we measure directly with the apparent optical depth method. We note that a larger velocity window of ± 20 km s^{-1} would result in a slightly higher (0.06 dex) column density limit. We also note that the strongest Si II line at λ1260 is overwhelmingly contaminated by intervening H I absorption and we use the Si II λ1526 line to set a limit for this ion.

3.3.4 Metallicity limits

Having established the H I column densities and limits to the ionic column densities of Si, C, and O, we may set upper limits to the heavy element abundances in the two LLSs. This requires, however, a careful consideration of the ionization state of the gas. Generally, the ionization state is assessed through a comparison of observed ionic column densities with ionization models. Absent the detection of *any* metal transitions, we must consider physically-motivated scenarios for the properties of the gas and thereby estimate limits to the metallicities.

In a perfectly neutral medium, all of the elements would be in their atomic state

(ignoring molecular formation) and one could estimate the metallicity directly from the observed column densities of the atomic transitions (e.g. H I C I, O I). In this extreme model one would recover $\log(\mathrm{O/H}) - \log(\mathrm{O/H})_\odot \equiv [\mathrm{O/H}] < -1.9(-1.4)$ and $[\mathrm{C/H}] < -2.1(-1.3)$ for LLS1134a (LLS0956B), which are the most conservative limits possible. In essentially all astrophysical environments, however, the gas is irradiated by local and external sources which photoionizes at least the outer layers of the cloud. Photons with energy $h\nu < 1\,\mathrm{Ryd}$ will photoionize all elements whose first ionization potential is below that of hydrogen (e.g. C, Si) and harder photons may ionized a majority of the H I gas and place the heavy elements in higher ionization states. Collisional processes, predominantly with electrons and hydrogen nuclei, may further ionize the medium for gas at $T > 10^4\,\mathrm{K}$.

Focusing on photoionization first, the population of galaxies and quasars at $z \sim 3$ is known to generate an extragalactic UV background (EUVB) which ionizes the majority of baryons in the universe (e.g. Haardt & Madau 1996, Rauch et al. 1997). For gas with the relatively low H I column densities of our LLSs, one predicts that the EUVB has substantially ionized the hydrogen gas and any heavy elements that are present. To quantitatively assess this process and its impact on the metallicity estimates, we have calculated a series of photoionization models using the CLOUDY software package (Ferland et al. 1998). We assumed that each LLS may be modeled as a series of layers with constant hydrogen density n_H, constant and low metallicity, and a plane-parallel geometry. We irradiated this gas with the Haardt and Madau EUVB model[3] (quasars+galaxies) (Haardt & Madau 2011) evaluated at the redshift of each LLS (e.g. $J_\nu = 2.45 \times 10^{-22}\,\mathrm{erg\,s^{-1}\,Hz^{-1}\,cm^{-2}}$ at 1 Ryd for $z = 3.4$) and considered gas with a range of $\mathrm{n_H}$ values. It is standard practice to describe a given photoionization model by the ionization parameter $U \equiv \Phi/(n_\mathrm{H} c)$ with Φ the flux of ionizing photons. Each model was restricted to have

[3]The results would be slightly different with a different shape for the radiation field, but the differences are much smaller than those related to varying the ionization parameter.

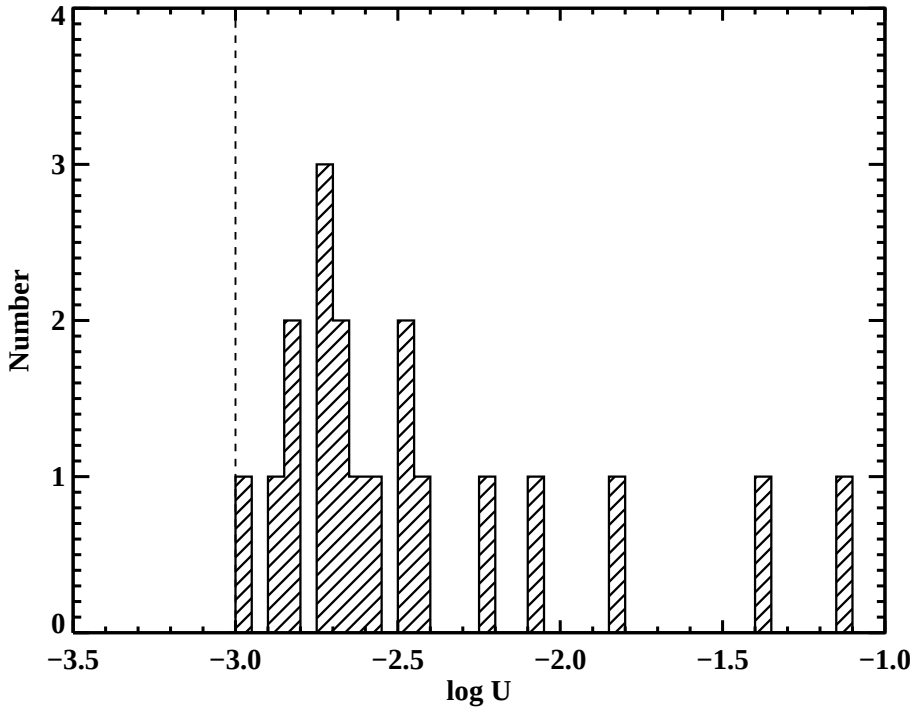

Figure 3.10: Distribution of the ionization parameter U for all the $z > 1.5$ LLSs reported to date in the literature. The value $U = 10^{-3}$ used in this analysis is marked with a dashed line.

a total n_{HI} column density consistent with the observations.

In Figure 3.11, we present the upper limits to the metallicity of the two LLSs as a function of n_H values. The figure indicates the limit imposed for the range of ions constrained by our observations. As a function of n_H, each curve shows the metallicity limit from a given ion according to the measured upper limit on its column density. The black symbols then trace the lowest metallicity imposed by the full set of measurements at each n_H value. In all cases we compare to the solar abundances of Asplund et al. (2009).

At high densities (low U values), the tightest limits are given by the lowest ionization stages of the heavy elements (Si^+, O^0, C^+). These set upper limits to the metallicity of $\lesssim 10^{-3} Z_\odot$. We emphasize, however, that such high densities are improbable. A gas with $n_H >$

$10^{-1}\,\mathrm{cm}^{-3}$ is predicted to have a neutral fraction $x_{\rm HI} \equiv n_{\rm HI}/n_{\rm H} > 0.1$, but all of the LLSs that have been studied to date exhibit substantial absorption from higher ionization states of Si, C, and O. When analyzed as photoionized gas, one derives ionization parameters $\log U = -3$ to -1 (Figure 3.10). At the respective redshift for the two LLSs, this implies a volume density $n_{\rm H} < 10^{-2}\,\mathrm{cm}^{-3}$ which corresponds to an overdensity relative to the mean baryon density of $\rho/\rho_{\rm mean} < 850$. As $\mathrm{n_H}$ decreases (and U increases), the heavy elements shift to higher ionization states and the limits on the gas metallicity become much more stringent. This is partly a result of tighter limits to the ionic column densities but it is primarily because our models are forced to match the observed $\mathrm{n_{HI}}$ value. At higher ionization parameter, the gas is more highly ionized and the total implied hydrogen column density is correspondingly larger ($N_{\rm H} = N_{\rm HI}/x_{\rm HI}$). Therefore, at fixed metallicity this implies higher total column densities for the heavy elements. For the ionization parameters typically measured for LLSs ($U > 10^{-3}$), we set an upper limit to the gas metallicity (based on Si) of $Z < 10^{-4.2}(10^{-3.8})Z_{\odot}$ for LLS1134a (LLS0956B). This is a conservative value and also the most physically-motivated limit to the metallicity.

If the intensity of the ionization field is much higher than the assumed EUVB, the lack of substantial Si IV and C IV absorption imposes a tight limit to the metallicity, as evident from Figure 3.11. Extreme scenarios in which most of the metals are in even higher ionization states are highly implausible since at $\log U \gg 1$ a neutral fraction $x_{\rm HI} \ll -6$ would imply a total hydrogen column density $N_{\rm H} \gg 10^{24}$ cm^{-2}. Further, if these two LLSs lie in proximity to a star forming galaxy, the ionization field is dominated by a softer spectrum. To explore this possibility, we have calculated a second set of CLOUDY models adding to the EUVB the contribution from a galaxy with star formation rate 100 $\mathrm{M_{\odot}\ yr^{-1}}$. We have further assumed that the escape fraction of ionizing radiation is $f_{\rm esc} = 0.1$ and that the LLSs are at 100 kpc from the star forming disks. Under these conditions, the ionization states of the metals do not differ substantially from our previous calculation. Indeed, at $N_{\rm HI} = 10^{18}$ cm^{-2}, the inferred

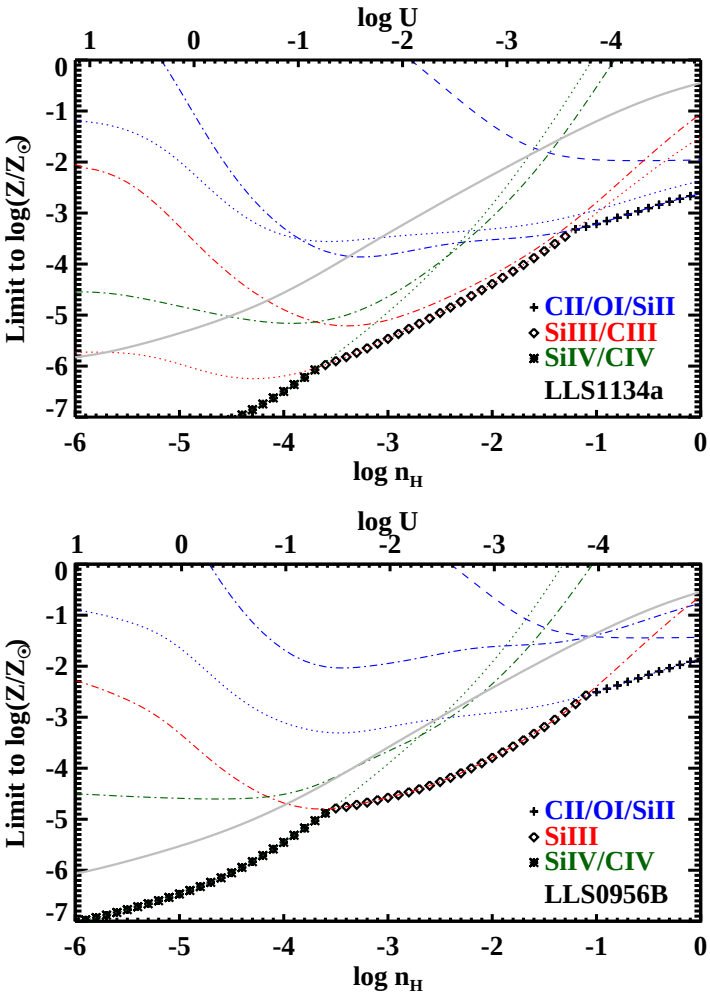

Figure 3.11: The two panels describe the metallicity limits to (top) LLS1134a and (bottom) LLS0956B as a function of the gas density $n_{\rm H}$, assuming the Haardt and Madau EUVB radiation field (Haardt & Madau 2011). Each curve shows the metallicity limit for a given ion according to the upper limit on its column density. Elements C, O, and Si are traced by dotted, dashed, and dash-dot lines respectively. The black symbols then trace the lowest metallicity imposed by the full set of measurements at each $\rm n_H$. The solid gray curves in the figure indicate the logarithm of the neutral fraction of the gas $x_{\rm HI} \equiv n_{\rm HI}/n_{\rm H}$.

metallicity for LLS1134a is $Z < 10^{-4.0}$. Therefore, we conclude that the assumed limits of $\sim 10^{-4}$ for LLS1134a and LLS0956B are representative of the underlying gas metallicity in different plausible scenarios.

For gas with $T > 2 \times 10^4$ K, collisional ionization is an important process for establishing the ionization state of the gas. If one assumes collisional ionization equilibrium, it is straightforward to set limits on the gas as a function of the gas temperature (Gnat & Sternberg 2007). For $T > 2.5 \times 10^4$ K, the limits are yet lower than those suggested by the photoionization modeling. At $T \sim 5 \times 10^4$ K, the lack of substantial Si III and C III constrains the metallicity at $Z < 10^{-6}(10^{-5})Z_\odot$ for LLS1134a (LLS0956B), while at higher temperatures, $T \sim 10^5$ K, the upper limits on the Si IV and C IV column densities correspond to $Z < 3 \times 10^{-7}(10^{-6})Z_\odot$ for LLS1134a (LLS0956B). At temperatures below $T \sim 2.5 \times 10^4$ K, we expect both collisional ionization and photoionization to contribute, with the latter dominating.

3.3.5 Comparison with other observations

To compare the limits on the metallicity for LLS1134a and LLS0956B with other observations, we search the literature for metallicity estimates of cosmic structures in the redshift interval $2 \lesssim z \lesssim 5$. The metal abundances in solar units as a function of the gas density are summarized in Figure 3.1. Since we cannot guarantee completeness in our search, this figure offers only a qualitative assessment on the interval of metallicity reported in one or multiple studies which include large samples. Here we provide a description of the different data shown in the figure.

The range of densities and metals for the Lyα forest is taken from Simcoe et al. (2004), who compile O VI absorption statistics at $z \sim 2.5$ using the line-fitting technique. These metallicities are consistent with other studies that are based on different analysis of the IGM (e.g. the pixel optical depth method Aguirre et al. 2008). Upper limits and statistical fluctuations

around or below $Z \sim 10^{-3}$ have been also reported at densities $\rho/\rho_{\rm mean} \sim 1$. These are represented in Figure 3.1 using a downward arrow.

For the LLSs (here loosely defined in the interval $\log N_{\rm HI} \sim 10^{16} - 10^{20.3}$ cm^{-2}), we search in the literature for published metallicity and densities between $z = 1.5 - 4.5$ (Prochter et al. 2010, Prochaska & Burles 1999, Kirkman et al. 2003, D'Odorico & Petitjean 2001, Prochaska et al. 2006, Prochaska 1999, O'Meara et al. 2001, Burles & Tytler 1998;?, Kirkman et al. 2000, Crighton et al. 2004, Levshakov et al. 2003a, Reimers & Vogel 1993, Sargent et al. 1990, Levshakov et al. 2003b). The range plotted in Figure 3.1 brackets these observations. For this class of objects we also present the logarithmic mean and standard deviation based on our (nearly) complete compilation. In addition to the metallicity, we gathered information on the ionization parameters, used to produce Figure 3.10. ISM metallicity in damped Lyα systems has been measured with both quasars and $\gamma-$ray bursts (GRBs). In both cases, we adopt a number density $0.05 - 10\,{\rm cm}^{-3}$ which is consistent with the densities for the ISM atomic phase (Spitzer 1978, Jorgenson et al. 2010). For the quasar sample, we report the range of metallicity between $z = 2 - 5$ from (Prochaska et al. 2003), extended at the lower end with the metal poor sample of Penprase et al. (2010). For the GRB sample we adopt instead the metallicities presented in Prochaska et al. (2007a).

Spectroscopy of high redshift galaxies provide abundances from unresolved H II regions in star forming galaxies. Here, we adopt the metallicity in two samples of $z \sim 2$ and $z \sim 3$ galaxies from Erb et al. (2006), Mannucci et al. (2009). We adopt a range of densities typical of nearby H II regions (Osterbrock 1989). Several studies have revealed super-solar metallicity in the broad line regions of quasars. Here, we adopt a compilation of metallicity from a large sample of $z > 3.7$ quasars and we choose a density interval typically measured in the broad line region (Osterbrock 1989).

Finally, we present the metallicity in Milky Way halo stars. Stellar metallicity is often

quoted using iron abundance, but the density of heavy elements is dominated by O, C, Ne, and other α-elements. We therefore consider the oxygen abundances presented in Fabbian et al. (2009), extended at low metallicity with the star reported by Caffau et al. (2011). Although the inferred metallicity at the lower end suffers from considerable observational uncertainties (Fabbian et al. 2009), values at $Z \sim 10^{-3} Z_{\odot}$ are common in large compilations (Suda et al. 2011) even for stars with extremely low iron abundances. In fact, only a single star has been reported at $Z \sim 10^{-4} Z_{\odot}$ (Caffau et al. 2011).

Chapter 4

Directly imaging damped Lyman-α galaxies at $z > 2$ – I. Methodology and first results

After gas accretion, the second key ingredient in models of galaxy formation is star formation. Understanding the conversion of gas into stars is a particularly challenging task at high redshifts, the principal limiting factor being the difficulty of observing molecular gas in emission or even the impossibility to probe the distribution of neutral hydrogen via 21−cm emission at cosmological distances. Thanks to recent instrumental upgrades at mm frequencies, however, the connection between molecular gas and star formation can now be explored even at $z > 2$ (Tacconi et al. 2010, Daddi et al. 2010). Conversely, the only possibility to characterize the relation between atomic hydrogen and star formation at cosmological distances is by studying damped Lyman-α systems that constitute the major repository of neutral hydrogen in the distant Universe. However, because of the bright background quasars that outshine faint nearby galaxies, previous studies (e.g. Wolfe & Chen 2006, Rafelski et al. 2011) were able to connect

the hydrogen properties of DLAs to the star forming properties of Lyman break galaxies only statistically. In this chapter, we present a reprint of the paper "Directly imaging damped Lyman-α galaxies at $z > 2$ – I. Methodology and first results" by Fumagalli et al. (2010b) in which we introduce a new imaging survey that, owing to a particular observational technique (O'Meara et al. 2006), is designed to directly connect the hydrogen properties of damped Lyman$-\alpha$ systems as measured in absorption to their local star formation rates inferred in emission.

4.1 Abstract

We present the methodology for, and the first results from, a new imaging programme aimed at identifying and characterizing the host galaxies of damped Lyman-α absorbers (DLAs) at $z \gtrsim 2$. We target quasar sightlines with multiple optically-thick H I absorbers and use the higher-redshift system as a "blocking filter" (via its Lyman-limit absorption) to eliminate all far-ultraviolet (FUV) emission from the quasar. This allows us to directly image the rest-frame FUV continuum emission of the lower-redshift DLA, without any quasar contamination and with no bias towards large impact parameters. We introduce a formalism based on galaxy number counts and Bayesian statistics with which we quantify the probability that a candidate is the DLA host galaxy. This method will allow the identification of a *bona fide* sample of DLAs that are too faint to be spectroscopically confirmed. The same formalism can be adopted to the study of other quasar absorption line systems (e.g. Mg II absorbers). We have applied this imaging technique to two QSO sightlines. For the $z \sim 2.69$ DLA towards J073149+285449, a galaxy with impact parameter $b = 1.54'' = 11.89$ h_{72}^{-1} kpc and implied star formation rate (SFR) of ~ 5 h_{72}^{-2} $M_\odot$ yr^{-1} is identified as the most reliable candidate. In the case of the $z \sim 2.92$ DLA towards J211444-005533, no likely host is found down to a 3σ SFR limit of 1.4 h_{72}^{-2} $M_\odot$ yr^{-1}. Studying the H I column density as a function of the impact parameter, including 6 DLAs with known

hosts from the literature, we find evidence that the observed H I distribution is more extended than what is generally predicted from numerical simulation.

4.2 Introduction

Absorption lines detected along the line of sight to quasi-stellar objects (QSOs) and gamma ray bursts (GRBs) can be used to glean the properties of the intergalactic medium (IGM) and the interstellar medium (ISM) at high redshift. Before the advent of large millimeter and radio arrays such as the Atacama Large Millimeter Array (ALMA) or the Square Kilometre Array (SKA), the only available way to characterize the physical properties of the different gas phases in the high redshift Universe is through the analysis of hydrogen, metal, and molecular absorption lines. Although the sizes of the regions explored through the background QSO beam are too narrow to provide a detailed picture of individual objects, large spectroscopic surveys of QSOs across the sky enable the study of the absorbers as a population. This can lead to profound insights on the gas properties in high-redshift galaxies, crucial to constrain models of galaxy formation and evolution.

One of the most well-studied classes of absorbers are the damped Lyα absorbers (DLAs). With an H I column density $N_{\rm HI} \geq 2 \times 10^{20}$ cm^{-2}, the DLAs contain most of the neutral gas in the Universe at $z \sim 3$ (O'Meara et al. 2007). Also, by being associated with high gas overdensities in the cosmic web, DLAs are intimately connected with galaxy formation at high redshifts.

Besides the actual identification of DLAs, absorption spectroscopy can provide detailed information about the H I column density distribution of the absorbers, their chemical composition and kinematics, as well as the physical state of the neutral hydrogen (see the review by Wolfe et al. 2005). As a result of several decades of observations of DLAs, the distribution

of neutral hydrogen in the Universe and its evolution with redshift is well constrained at high redshift (Prochaska et al. 2005, Prochaska & Wolfe 2009, Noterdaeme et al. 2009). Still, pencil beam surveys yield only a limited picture of the morphology of DLA galaxies. In turn, this limits the utility of the absorbers for studying galaxy assembly and evolution.

While at low redshift ($z < 1$), DLAs are clearly associated with galaxies (e.g. Zwaan et al. 2005a; and references therein), the nature of high redshift DLAs is still uncertain. Since their discovery (Wolfe et al. 1986), the absorbers have often been associated with massive disks, as suggested by velocity profiles of low-ion metal transitions (Prochaska & Wolfe 1997) and consistent with recent findings that massive thick disks with typical rotational velocities up to 200 km/s are already in place at redshift $z \sim 2 - 3$ (Genzel et al. 2006, Förster Schreiber et al. 2009). However, the abundance patterns in DLAs indicate star formation histories more similar to those of dwarf irregular galaxies (Dessauges-Zavadsky, Calura, Prochaska et al. 2007), while the elusive nature of the DLA galaxies hints towards a population of low surface-brightness systems.

From a theoretical point of view, smoothed particle hydrodynamic (SPH) simulations which include gas physics are able to reproduce most of the observed DLA properties within a cold dark matter (CDM) formulation (see, however, Jedamzik & Prochaska 1998). Although the results may depend on the treatment of feedback and winds, there is general agreement that the major contribution to the DLA cross-section at $z \sim 3$ comes from low- and intermediate-mass halos with $10^9 < M_{\rm vir}/M_\odot < 10^{12}$ (see also Barnes & Haehnelt 2009). This is consistent with the value $M_{\rm vir} = 10^{11.2} M_\odot$ inferred by Cooke et al. (2006) from the clustering of DLAs and Lyman-break galaxies (LBGs). Nevertheless, the debate around DLAs has not yet been settled. In fact, simulations tend to predict small impact parameters, suggesting that DLAs are more compact at high redshifts than modern disk galaxies. But this causes simulations to under-predict the observed rate of incidence (e.g. Nagamine et al. 2007) or the number of high

velocity absorbers (Pontzen et al. 2008). Several mechanisms such as tidal-streams, outflows (e.g. Schaye 2001) or filamentary structures (e.g. Razoumov et al. 2006) and cold flows penetrating inside massive halos (Kereš et al. 2005, Dekel et al. 2009b) can provide a larger cross-section for DLA gas. More quantitative analysis of adaptive mash refinement (AMR) simulations are ongoing to understand if gas overdensities inside these more extended structures can reproduce the spectrum of kinematics observed in DLAs, as well as the incidence of the absorbers.

To identify which one, or which combination, of the above scenarios applies to DLAs requires direct imaging of the galaxies responsible for the absorption. Unfortunately, this task is particularly difficult at optical wavelengths due to the bright emission of the background quasar. In the past years, several attempts have been made in this direction[1], typically by inspecting the residual images after subtracting out the quasar light. However, the galaxy counterparts of these absorbers are expected to be faint and probably at low impact parameters (e.g. Wolfe & Chen 2006, Nagamine et al. 2007). Therefore, imperfections of the quasar subtraction are a challenge to such studies (see Kulkarni et al. 2000; 2001). As a result, only six spectroscopically-confirmed galaxy counterparts are currently known at $z > 1.9$ (Møller & Warren 1993, Djorgovski et al. 1996, Fynbo et al. 1999, Møller et al. 2002; 2004).

To overcome these limitations, new techniques are being explored. Surveys based on adaptive optics and improved modelling of the QSO point spread function (PSF) can minimise the impact of the quasar light on nearby objects, although some regions at very small impact parameters may still not be accessible. Narrow-band images from integral field unit (IFU) observations have the great advantage of providing both spatial and redshift information at the same time. Unfortunately, current instruments do not provide very high sensitivity at the short wavelengths needed to detect the Lyα line at $z \sim 2$ (see Christensen et al. 2007). A very promising technique to image high-z absorbers was proposed by O'Meara et al. (2006),

[1]See Appendix 4.10.2 for a review of previous studies aimed at identifying DLA galaxies.

who considered imaging Mg II absorbers at $z \sim 2$. The basic idea, presented in more detail in section 4.3, consists of imaging QSO sightlines with two known high column density absorbers. The higher-redshift absorber can then act as a natural filter to block the quasar light, so that the rest-frame far-ultraviolet (FUV) emission of the lower-redshift DLA can be detected without any contamination from the quasar.

This paper, the first of a series, presents initial results from a new survey to image DLAs at $z = 2 - 3$, using the above technique. In section 4.3, we discuss the target selection criteria, in section 4.4, we describe the observations of two quasar fields, the data reduction procedure, and our results, while, in sections 4.5 and 4.6, we focus on different methods to identify the galaxy counterparts. Analysis and discussion follow in sections 4.7 and 4.8, while section 4.9 summarizes our present results and considers prospects for the future. We adopt a Λ-cold-dark-matter (ΛCDM) cosmology throughout this paper, with $\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$, and $H_0 = 72\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$. All lengths are proper distances unless otherwise stated. Physical quantities are computed including the Hubble constant, in units of $h_{72} = 0.72$.

4.3 Survey design

The selection criteria for our targets are based on an updated version of the O'Meara et al. (2006) method, used to eliminate quasar contamination. We search among all the known QSOs with a foreground DLA in the Sloan Digital Sky Survey (SDSS) that also harbour a higher-redshift Lyman-limit system (LLS)[2]. By requiring $N_{\rm HI} > 10^{18}$ cm^{-2} for the LLS, we only include absorbers that are highly optically thick ($\tau > 10$) to Lyman continuum photons. This configuration of two absorbers allows us to use the higher-redshift absorber to completely block the quasar light, allowing the FUV emission of the lower-redshift DLA to be imaged without

[2]We refer to the second absorber as an LLS to make clear the distinction with the target DLA, at a lower redshift, that is to be imaged. However, the higher-redshift absorber can also be a DLA.

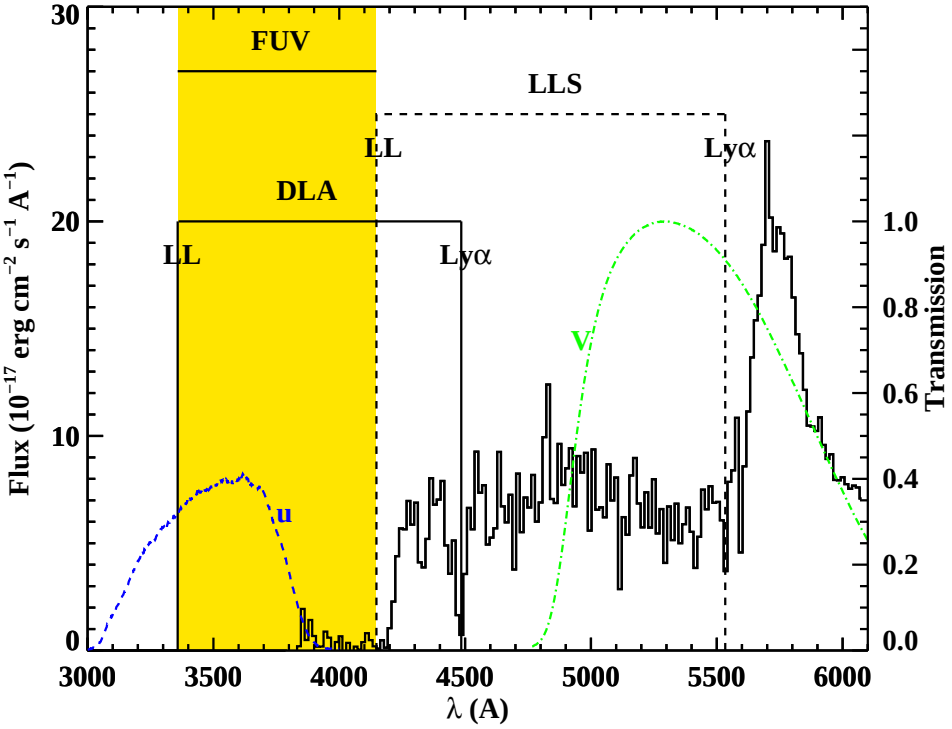

Figure 4.1: The SDSS spectrum of QSO J073149+285449 (black histogram). Also labeled are the Lyα and Lyman-limit wavelengths of two absorbing systems along the sightline: (i) a Lyman-limit system (LLS) at $z_{lls} = 3.55$ and (ii) a DLA at $z_{dla} = 2.69$. The yellow shaded region corresponds to the wavelength range where the rest-frame FUV emission from the lower-z DLA can be detected without quasar contamination. The u-band (blue dashed line) and V-band (green dash-dotted line) LRIS filter transmission curves are overplotted to show that the quasar light is fully absorbed by the LLS in the u-band image.

any quasar contamination or source confusion from the QSO host galaxy.

An example is provided in Figure 4.1, where we show the SDSS spectrum of the QSO J073149+285449. For illustrative purposes, the LRIS[3] u and V filter transmission curves are superimposed with blue dashed and green dash-dotted lines, respectively. The quasar spectrum exhibits an LLS at $z_{lls} = 3.55$ with a corresponding Lyα absorption line at ~ 5500Å and a Lyman limit (LL) at $\lambda_{lls}^{\rm LL} = 912$ Å$\times(1+z_{lls}) \sim 4150$Å. In addition, Ly$\alpha$ absorption from a lower redshift DLA can be seen at ~ 4500Å, with an associated Lyman limit at $\lambda_{dla}^{\rm LL} = 912$ Å$\times(1+z_{dla}) \sim 3400$Å. The higher-redshift LLS entirely absorbs the quasar light at $\lambda < \lambda_{lls}^{\rm LL}$, as seen in the spectrum. This allows the lower-redshift DLA to be imaged in filters covering wavelengths blueward of $\lambda_{lls}^{\rm LL}$.

A comparison between the u-band and R-band images of two such fields, containing the quasars J211444-005533 (top) and J073149+285449 (bottom), is shown in Figure 4.2. The QSOs are visible only in the R-band images (the left panels of the figure), while they are fully absorbed by the higher-redshift LLSs in the u-band images (the right panels). This allows the detection of faint lower-redshift galaxies at all impact parameters. It is useful to note that, even if the LLS is associated with a star-forming galaxy, it is unlikely to be visible in the u filter unless the galaxy has a very high escape fraction. Moreover, the UV light from the LLS will eventually recover from the absorption at $\lambda \ll \lambda_{lls}^{\rm LL}$, but without a significant contribution in the u-band. This prevents confusion for the identification of the lower-z DLA. The LLS could be detected in the V- or R-band imaging as a Lyman-break galaxy, unless it is projected onto the quasar.

Next, since the DLA has its own Lyman limit, its FUV emission can be detected only

[3]Low Resolution Imaging Spectrometer at Keck I (Oke et al. 1995).

in the wavelength interval

$$\lambda_{dla}^{\rm LL} \equiv 912\ \mathring{A} \times (1 + z_{dla}) < \lambda < \lambda_{lls}^{\rm LL} \equiv 912\ \mathring{A} \times (1 + z_{lls})\,, \tag{4.1}$$

highlighted with a yellow shaded area in Figure 4.1. Therefore, it is strategic to impose the selection criterion that $z_{dla} \ll z_{lls}$, so as to maximise the emission from the DLA in the u-band filter. Conversely, a lower limit on the DLA redshift is imposed by the condition 1215 Å$\times(1 + z_{dla}) > 912$ Å$\times(1 + z_{lls})$, i.e. the DLA Lyα line has to be visible in the spectrum. The latter improves upon the original selection criterion of O'Meara et al. (2006), who relied on strong metal lines (e.g. Mg II, Fe II, etc) to infer the presence of a high column-density absorber (see also Christensen et al. 2009). While strong Mg II absorbers and DLAs are often considered highly overlapping populations (Rao et al. 1995), only $\sim 35\%$ of strong Mg II absorbers have been found to have a column density above the DLA limit (Rao et al. 2006). An additional advantage of our approach is that the H I column density for our targets is directly measurable from the Lyα absorption line, implying that it is possible to determine the metallicity and relative abundances in the absorber's ISM along the QSO sightline. The downside is that we impose another constraint on the redshift separation of the two absorbers which limits the number of possible targets. Finally, two additional conditions set the absolute redshift space that we can probe with our survey using current technology. An upper limit at $z_{dla} \sim 3.5$ is imposed by the absorption from the intergalactic medium (IGM). In fact, at higher redshifts the blanketing effect of the IGM starts affecting the emission from the DLA galaxy, lowering the chance of a detection. Conversely, a lower limit at $z_{dla} \sim 2.1$ is imposed by the target selection using SDSS and, more generally, by the use of optical rather than UV facilities. We note, finally, that the short wavelength imaging can be carried out with either ground- or space-based facilities.

At first, the requirement that two absorbers should lie in a narrow range of redshifts

along a single sightline may suggest that we will be able to target only a few systems in this particular spatial configuration. However, among the ~ 1000 DLAs known at $z \gtrsim 2.1$ from the SDSS (DR5; Prochaska et al. 2005), ~ 140 sightlines meet our selection criteria. Therefore, the proposed technique is a promising way of obtaining a large sample of DLAs for a statistical study of the emission properties of the host galaxies. Note that it is important to restrict the wavelength range that is imaged to the region between the Lyman limits of the two DLAs, to minimize both the leakage from the QSO and the sky emission at $\lambda \lesssim \lambda_{dla}^{\rm LL}$. Using a tunable medium-band filter would be ideal for this project, but such filters are not typically available on large telescopes. To isolate a first sample of high-priority targets, we require that the broad-band filters that are currently available overlap with the FUV visibility window defined by Eq. (4.1). With these additional constraints, we have selected a sample of ~ 40 sightlines, 20 of which will be imaged with Wide Field Camera-3 (WFC3) on the Hubble Space Telescope (HST)[4], and ~ 20 with ground-based facilities.

In summary, with this survey we aim to increase the number of known host galaxies of high-z DLAs, over a wide range of both redshift ($z = 2 - 3.5$) and H I column density ($N_{\rm HI}= 2\times10^{20} - 7\times10^{21}$ cm^{-2}). Since any bias towards large impact parameters is completely removed, even non-detections of DLA emission will provide interesting constraints on the star formation rates (SFRs) of the absorbers. While such non-detections might have been attributed to the quasar glare in previous studies, our technique will yield robust upper limits on the DLA luminosities.

[4]The HST-WFC3 observations have been scheduled for the ongoing Cycle 17, proposal ID 11595.

4.4 Observations

We have applied the technique described in the preceding section to two QSOs, J211444-005533 and J073149+285449, each with two high H I column density absorbers along the sight-line. Details concerning the quasars and the absorbers are provided in Table 4.1. The last column lists the fraction of the u-band filter transmission $g(\lambda)$ that covers the FUV window $\lambda_{dla}^{\rm LL} < \lambda < \lambda_{lls}^{\rm LL}$ in which the DLA can be imaged:

$$f(FUV) = \frac{\int_{\lambda_{dla}^{LL}}^{\lambda_{lls}^{LL}} g(\lambda)\mathrm{d}\lambda}{\int_{0}^{\infty} g(\lambda)\mathrm{d}\lambda} \ . \tag{4.2}$$

In the case of J211444-005533, the "blocking" absorber is a system at $z \sim 3.44$, associated with the quasar, while the target absorber is a super Lyman-limit system (SLLS; or a sub-DLA[5]), at $z \sim 2.92$. Conversely, for J073149+285449, the blocking absorber and the intervening DLA are at $z \sim 3.55$ and $z \sim 2.69$, respectively.

Imaging of the fields of J211444-005533 and J073149+285449 was obtained at Keck I using LRIS. The first field was observed in October 2008, during a photometric night, and the second field in January 2009, during a stable but non-photometric night. A set of short exposures for J073149+285449 were subsequently acquired in a photometric night for flux calibration. The blue side of LRIS is equipped with a 2×2K×4K back-side-illuminated Marconi CCD with a plate scale of 0.135 $''$ pix^{-1}. Before June 2009, a 2K×4K front-side-illuminated Textronic CCD with a plate scale of 0.211 $''$ pix^{-1} was in operation on the red side.

[5] J211444-005533 has been included in our sample since the measured value of column density together with the associated error places this object at the edge of the DLA classification.

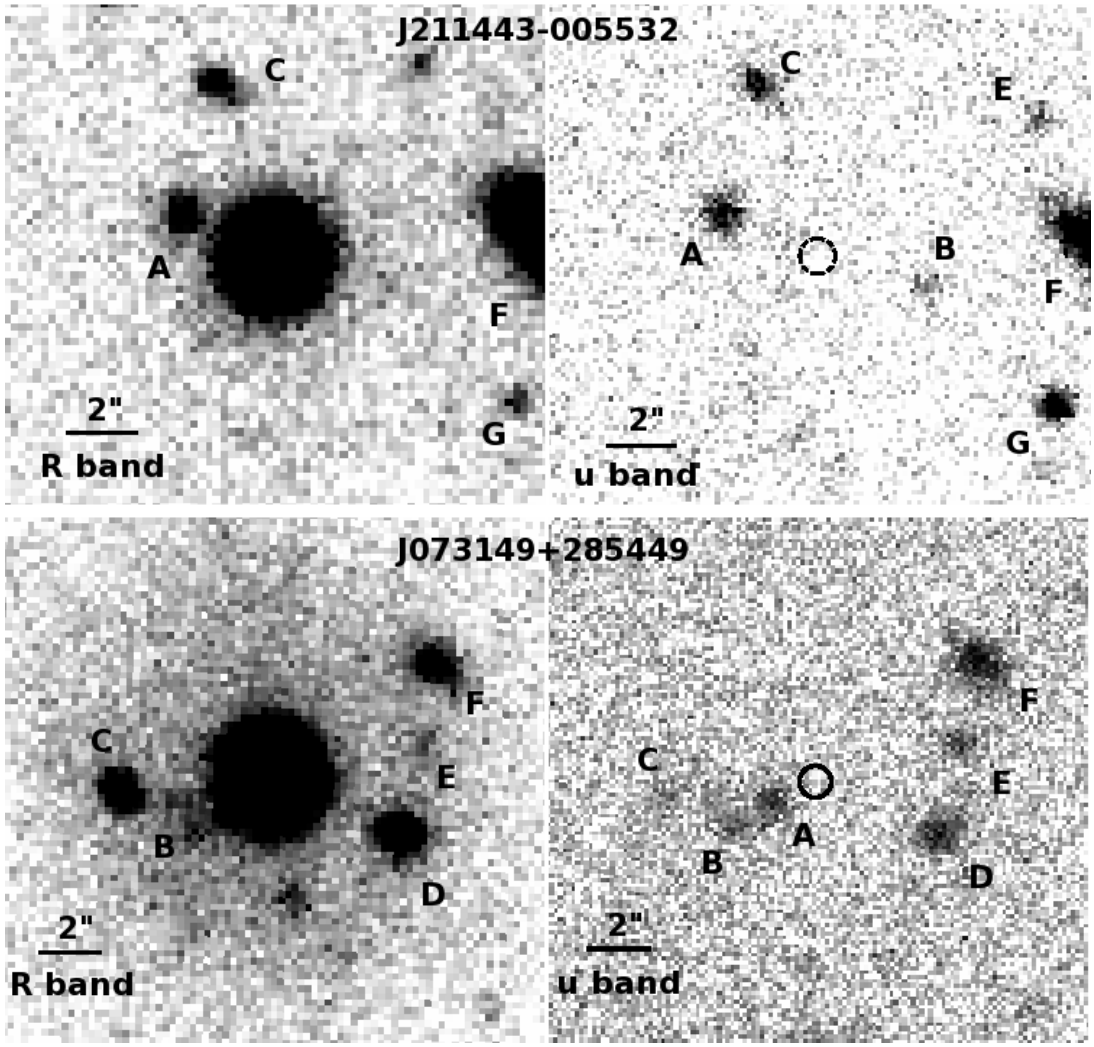

Figure 4.2: Keck u- and R-band imaging of the fields J211444-005533 (top) and J073149+285449 (bottom). The quasars are visible only in the R-band images (left panels) because the intervening LLS completely absorbs the light in the u-band (right panels). Therefore, faint galaxies in the foreground of the LLS can be detected even if they are spatially coincident with the quasar. Galaxies detected in the u band are labeled as in Table A4.6, while the QSO position is marked with a circle of $0.5''$ in radius. The solid lines are $2''$ long (~ 15 h_{72}^{-1} kpc at $z = 3$). Galaxy F in the field J211444-005533 is the only object visible in the SDSS images.

Table 4.1: A summary of the properties of the absorbers detected towards J211444-005533 and J073149+285449, and of the quasars themselves. Quasar magnitudes in the r band are from the SDSS. Also listed are the observed wavelengths of the Lyman limit of the LLS ($\lambda_{lls}^{\rm LL}$), and the fractions of the filter covered in the FUV window, as defined in the main text.

DLA	z_{qso}	r_{qso} (mag.)	z_{lls}	$\lambda_{lls}^{\rm LL}$ (Å)	z_{dla}	log $N_{\rm HI}$ (cm^{-2})	f(FUV)
J211444-005533	3.44	18.67	3.44	4052	2.92	20.20	0.38
J073149+285449	3.68	18.47	3.55	4152	2.69	20.55	0.73

4.4.1 Imaging

We acquired multiple exposures for each target, dithering $\sim 15''$ to remove CCD defects in the final image. A summary of the observations is in Table 4.2. By splitting the incoming light through a dichroic mirror (50% transmission at 4874Å), R-, V- and I-band images were obtained for J211444-005533, simultaneous with the u-band exposures. During the observations of J073149+285449, water vapor condensed on the window of the red-side camera, producing a halo around the quasar (see bottom-left panel of Figure 4.2). For this target, besides the u-band image, we hence only acquired R- and V-band images, which have limited value. Observations were taken close to the meridian in order to minimise the atmospheric extinction. Seeing conditions were good (FWHM$\sim 0.6'' - 0.8''$ in the u band). The data were reduced following standard procedures. After the bias subtraction, we applied twilight flats and then averaged background-subtracted exposures after scaling them to a common zero. A weight proportional to the background variance was adopted for the stacking.

Photometric calibrations were obtained by observing multiple photometric standard stars in the fields PG2213−006 and PG0918+029 (Landolt 1992). A photometric zero-point in AB magnitude was fitted together with a color term, assuming fixed air mass coefficients typical for the atmosphere in Mauna Kea (0.41, 0.12, 0.11 and 0.07 for u, V, R, and I; Cooke et al.

2005). For galaxies with fluxes affected by the quasar emission in all filters besides u-band, we set the color term to zero, assuming a flat continuum typical of star-forming galaxies. Uncertainties on the final zero-point are between ~ 0.05 and ~ 0.02 mag. for J211444-005533, while between ~ 0.07 and ~ 0.04 mag. for J073149+285449. The higher uncertainty for J073149+285449 is due to the intermediate step required to extrapolate the zero-point from shallow exposures acquired in photometric conditions. Corrections for Galactic extinction (Table 4.2) are computed from the far-IR dust map of Schlegel et al. (1998). The extinction A in each filter is computed as $A(u) = 4.8E(B-V)$, $A(V) = 3.1E(B-V)$, $A(R) = 2.3E(B-V)$, and $A(I) = 1.5E(B-V)$ (Cardelli et al. 1989). Under good seeing conditions, a total Keck-LRIS exposure time of $\sim$ 90 min. enables a depth of $\sim$ 29 mag. at 1σ for a $1''$ aperture in the u-band images. This sensitivity allows the detection of a star formation rate of ~ 1.5 h_{72}^{-2} $M_{\odot}$ yr^{-1}at 3σ significance for a $z = 3$ target, once we correct for IGM absorption (see Sect. 4.7). Exposure times and depths in each filter are listed in Table 4.2.

Table 4.2: Log book of the imaging observations taken at Keck-LRIS.

Field	R.A. (J2000)	Dec. (J2000)	UT Date	Filter
J211444-005533	21:14:43.9	-00:55:32.7	2008 Oct, 2nd	u
				V
				R
				I
J073149+285449	07:31:49.5	+28:54:48.7	2009 Jan, 28th	u
				V
				R

Field	Exp. Time (s)	FWHM ($''$)	1σ depth (mag in $1''$ ap.)	E(B-V) (mag)
J211444-005533	6×900	0.6	29.20	0.062
	6×220	0.6	28.22	
	6×220	0.6	27.99	
	6×245	0.6	27.59	
J073149+285449	6×900	0.7	28.88	0.055
	6×360	0.8	27.83	
	6×360	0.7	27.60	

4.4.2 Photometry

Candidate host galaxies were selected from the u-band images using the SExtractor package (Bertin & Arnouts 1996). The detection threshold was set to 1.4σ with a minimum area of 3 pixels; these parameters force the inclusion of faint sources. We include in the final catalogue only galaxies within a projected angular distance of $b < 12''$ from the quasar (corresponding to a proper distance of ~ 90 h_{72}^{-1} kpc at $z = 3$). This search area is slightly larger than the maximum impact parameter of an absorber ($\sim 10''$), as inferred from absorption line statistics (Storrie-Lombardi & Wolfe 2000). Since the region under consideration is small, we can inspect the segmentation maps to clean the catalogue of spurious detections or to include undetected sources, if any.

Integrated magnitudes are computed within Kron-like elliptical apertures. The background is subtracted locally, measuring the sky mean value in a square box 40 pixels on a side, centered on the target. All the pixels flagged as belonging to an object are first masked, and the sky variance is added to the Poisson error from the source to compute the uncertainty on the flux measurement. The final uncertainty also includes the error on the photometric calibration. To test the accuracy of the photometry, we have simulated Keck u-band observations for different seeing conditions with the software SkyMake (Bertin 2009). For good seeing ($0.6''$), we are able, on average, to recover $\gtrsim 95\%$ of the total flux at 1σ down to $u = 25$ mag, while, for the faintest magnitudes ($u \gtrsim 26$ mag), this fraction approaches $\sim 90\%$. In all cases, the total flux is fully recovered within 2σ significance. For worse seeing conditions ($1.0''$), the fraction of recovered flux drops slightly, as expected. Due to the increase in the uncertainty, the fraction of flux recovered at 1σ remains constant.

An additional source of uncertainty comes from possible leakage of the quasar flux.

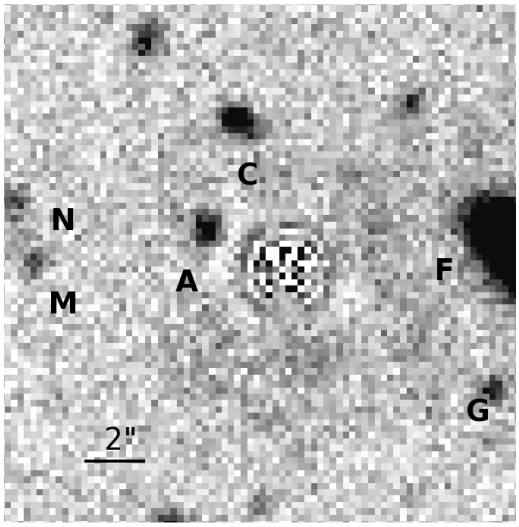

Figure 4.3: R-band image of J211444-005533, after quasar subtraction using standard PSF modelling techniques. Galaxies detected in the u band are labeled as in Table 4.6.

Although $\tau_{lls} \gg 1$ for the H I column densities of the higher-z LLSs, we conservatively test for any possible contamination from the quasar by comparing the surface brightness in a box centered on the quasar region with a local sky determination. The difference in surface brightness normalised to the sky variance is $\Delta_\mu/\sigma = -0.01$ for J211444-005533 and $\Delta_\mu/\sigma = -0.03$ for J073149+285449. Since these discrepancies are within a few percent of the sky variance in both cases, we conclude that the intervening LLSs are effective in fully blocking the light from the background quasars.

Table 4.6 in the appendix provides photometric information for objects detected at $S/N > 3$. However, we will conservatively consider only targets with $S/N > 5$ in the u band to be candidates for the DLA counterparts. The photometry in the R, V, and I filters is mainly intended to provide colors for the photometric redshift analysis, rather than an accurate determination of the total flux. To alleviate color gradient effects and seeing differences, we compute the half-light radius (r_{hl}) for our targets on a white image, produced by stacking the

R-, V-, and I-band images, where available. For all the galaxies detected in the u band, we compute colors in circular apertures, multiples of r_{hl}. Some candidates lie at very small impact parameters to the quasar sightline and their colors need to be corrected for quasar contamination. Therefore, we model and subtract the quasar light-profile by fitting a 4^{th}-order b-spline model (See Appendix A of Bolton et al. 2006). An example of the residual image in the R band is presented in Figure 4.3 for J211444-005533. Although the result is quite satisfactory, the residuals may still affect the photometry and we choose to use only colors which are stable to the quasar subtraction. As already noted, the red-side images for the field of J073149+285449 were affected by instrumental problems; since we are not able to model the scattered light, we do not present colors for this field.

4.4.3 Impact parameter

The impact parameter b is defined as the proper distance at the absorber redshift between the line of sight to the quasar and the center of the absorbing galaxy, the latter computed as the first moment of the light distribution. Since the quasar is completely absorbed in the u band, we transfer the quasar position from the R-band image to the u-band image, using accurate relative astrometry. This is done by first fitting an astrometric solution over the R-band image, using stars with known positions. We then fit a second astrometric solution to the u-band image, using more than five reference objects, whose positions are extracted from the R-band image and selected to be within $\sim 20''$ from the quasar. Using this procedure, we achieve a high accuracy for the distances of objects close to the QSO sightline, better than that obtained from a single astrometric solution. The typical errors on the angular separation from this procedure are $0.05''$ for J211444-005533 and $0.07''$ for J073149+285449 (corresponding to ~ 0.4 h_{72}^{-1} kpc at $z = 3$). Once the quasar position is known in the u band, we compute the projected quasar-galaxy angular separation (b_{as}) for each candidate host galaxy. The angular distance is then

converted into a physical separation b_p, assuming comoving distances of $D_c = 6102$ h_{72}^{-1} Mpc for J211444-005533 and $D_c = 5868$ h_{72}^{-1} Mpc for J073149+285449. The impact parameters obtained for the different candidates are listed in Table 4.3. The lowest impact parameter for candidates in the J073149+285449 field is $\sim 1.54''$ (i.e. 11.89 h_{72}^{-1} kpc at $z = 2.686$, the DLA redshift), while that for systems in the J211444-005533 field is $\sim 2.86''$ (i.e. 21.61 h_{72}^{-1} kpc at $z = 2.919$).

4.5 Identification of the DLA hosts

The u-band images described in the previous section reveal a number of candidates for the DLA host galaxy within an angular distance of $\sim 12''$ from the quasar. We wish to identify which, if any, of these is responsible for the observed damped absorption line. In studying absorption line systems, the impact parameter to the quasar sightline is often used to identify host candidates, following the general rule that the nearest object to the line of sight is likely to be the galaxy that causes the absorption. However, as will be shown later, the number of interlopers increases significantly in very deep searches; for this reason, a more quantitative treatment is needed. In this section, we discuss the two most straightforward methods to confirm that the DLA indeed arises in one of the candidate host galaxies. A more indirect, statistical approach to quantify the relative probability that one of the candidates gives rise to the DLA is described and discussed in the next section.

4.5.1 Spectroscopy

The only way to confirm a galaxy-absorber association for each system is through a spectroscopic detection of the galaxy, with an emission/absorption redshift consistent with the redshift of the DLA. For targets close to or aligned with the quasar, detecting Lyα emission in the DLA trough is a simple way to measure the redshift of the host galaxy. The quasar

light is blocked by the damped Lyα absorption, and one can hence search for Lyα emission from the same redshift with impunity (e.g. Møller et al. 2004). Previous searches for Lyα emission were mostly limited by lack of knowledge of the location of the star-forming regions in the host galaxy, due to which it was not clear where to place (and how to orient) the slit for a spectroscopic search. This meant that a non-detection of Lyα emission might arise simply because the brightest regions of the host galaxy were not covered by the chosen position and orientation of the slit. Crucially, our survey will directly yield the positions of the candidate host galaxies, allowing follow-up spectroscopic studies to correctly position slits on all candidates. Note that the detection of other spectral lines at optical wavelengths is likely to be affected by the bright quasar continuum. However, the Hα transition, redshifted into the near-IR waveband for DLAs at $z \gtrsim 2$, is the other plausible transition by which the galaxy redshift can be measured.

Beside Lyα (or Hα) emission, the technique adopted here allows an alternative route to confirm or at least constrain the galaxy redshift. Figure 4.1 shows that the QSO contamination disappears blueward of the Lyman break of the higher-redshift LLS; any continuum detected in this part of spectrum comes only from foreground objects that are in the slit. Therefore, we can establish the redshift of the candidate by identifying other absorption features such as metal lines. Furthermore, a less precise but still useful redshift determination can be obtained by searching for a signature of the galaxy Lyman limit, if visible redward to the atmospheric cut-off.

A limitation of the above spectroscopic methods is that dust extinction can suppress the Lyα or FUV emission. Also, our poor knowledge on the escape fraction of Lyα photons makes it very difficult to estimate the expected Lyα flux at a given UV luminosity (see for example Matsuda et al. 2004). In addition, the detection of the galaxy continuum in spectra can only be obtained within a reasonable integration time ($\sim 1-2$ hours at a 10m-class telescope) for targets brighter than 25 or 26 magnitudes. As discussed in the introduction, simulations

(and some observational studies in the literature) suggest that DLAs may be associated with even fainter objects. Finally, the Hα line is only observable from ground-based facilities from a narrow redshift range. For these reasons, it may not be possible to spectroscopically confirm all candidate host galaxies and different approaches are required.

Table 4.3: Projected angular and physical distances from the QSO sightline for each candidate host galaxy detected in the u-band images. Also quoted is the frequentist probability of the candidate's being an interloper.

ID	b_{as} (")	b_p (kpc)	P_f	ID	b_{as} (")	b_p (kpc)	P_f
	J211444-005533				J073149+285449		
A	2.86	21.61	0.19	A	1.54	11.89	0.06
B	3.27	24.69	0.56	B	2.87	22.14	0.28
C	5.10	38.50	0.51	C	4.54	35.01	0.82
D	5.07	38.30	0.97	D	4.33	33.45	0.29
E	7.46	56.34	0.98	E	4.74	36.59	0.62
F	8.67	65.43	0.08	F	6.33	48.86	0.38
G	7.93	59.85	0.82	G	10.46	80.76	1.00
H	10.25	77.40	0.82	H	12.69	97.92	0.95
I	9.61	72.52	0.58	I	12.38	95.55	0.18
L	8.41	63.46	1.00	L	8.59	66.30	0.98
M	8.35	63.07	0.78	M	9.92	76.56	1.00
N	9.15	69.07	0.89	N	12.69	97.94	1.00
O	5.12	38.64	0.99	-	-	-	-
P	12.04	90.87	1.00	-	-	-	-
Q	11.89	89.79	1.00	-	-	-	-
R	12.03	90.78	1.00	-	-	-	-
S	12.95	97.75	0.76	-	-	-	-
T	12.15	91.73	1.00	-	-	-	-

4.5.2 Photometric redshift

A second method to determine the redshifts of the candidate host galaxies is via a photometric redshift ("photo-z") estimate. The advantage of this technique over spectroscopy is that it can also be used for faint galaxies. Unfortunately, there are two main issues that affect the photo-z analysis. First, quasar contamination does not allow a robust color estimate

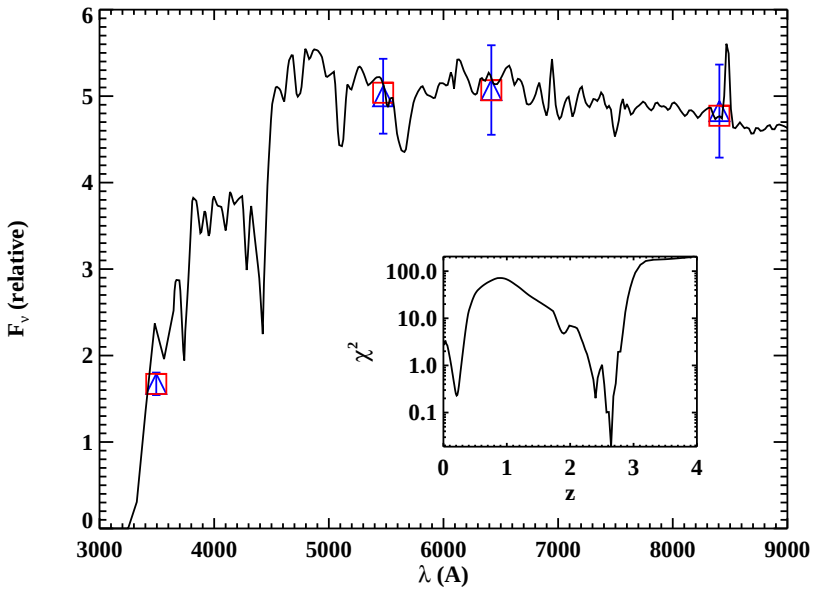

Figure 4.4: Photometric redshift for galaxy J211444-005533-C: SED, synthetic fluxes (red square) and observed fluxes (blue triangles with errors). The inset shows the χ^2 distribution as a function of galaxy redshift. The fitted redshift ($z = 2.65$) suggests that this galaxy is not associated with the DLA at $z_{dla} = 2.92$, but we cannot rule out the galaxy-absorber correspondence at $> 3\sigma$ C.L.

for targets at low impact parameter, i.e. those more likely to be associated with the DLA (see below). For this reason, the photo-z method can only be used to estimate the redshift for targets at large projected distances from the quasar. Second, photometry in 4 optical filters covers only a narrow range of the blue part of the spectral energy distribution (SED) of a galaxy. This significantly increases the number of catastrophic outliers (e.g. Hildebrandt et al. 2008), making the results less reliable. It is worth mentioning that one can try to constrain the redshift through photometry by combining our ground-based u-band imaging with HST UV observations in narrow- or medium-band filters.

As noted earlier, instrumental problems caused colors for the galaxies in the field of J073149+285449 to be contaminated by scattered light from the quasar, making these colors unreliable for a photo-z analysis. For the J211444-005533 field, we have computed photometric

redshifts for the candidate host galaxies using the `eazy` code (Brammer et al. 2008) and an SED template library from Grazian et al. (2006). The fit was performed on a grid of redshifts ranging from 0.01 to 4 with a resolution $\Delta z = 0.01$, including the effects of IGM absorption. Although the code allows linear combinations of SEDs, we used individual templates, without priors on the galaxy magnitude. We initially focused on targets A and C in the J211444-005533 field since these are the two galaxies with lowest impact parameters to the quasar line of sight. For J211444-005533-C, the best-fit redshift is $z = 2.65$. The SED, the synthetic fluxes (red square), and the observed fluxes (blue triangles) are shown in Figure 4.4, with the χ^2 distribution as a function of redshift displayed in the inset. No other significant relative minima are found besides the two in the same redshift interval. Conversely, for target A, the best fit is at $z = 2.50$, but a second minimum is found at $z \sim 0.2$, making this redshift determination less secure. Unfortunately, the limited number of available filters means that neither redshift can be constrained at a high confidence level (C.L.). In fact, although J211444-005533-A and J211444-005533-C appear to be located at lower redshifts than the DLA ($z_{abs} = 2.919$), we cannot rule out the galaxy-absorber correspondence at $> 3\sigma$ C.L. for either candidate; this illustrates the problems with the photo-z approach, and emphasizes the need for spectroscopic confirmation. Among the other galaxies with color determinations in the field of J211444-005533, we do not find any significant DLA candidates using the photo-z approach.

4.6 Statistical approaches

As discussed, spectroscopy is required to securely identify the host galaxies of DLAs. However, besides being an expensive observational task, it may even be unsuccessful in some cases. Statistical approaches to quantify the probability that a given galaxy is associated with a DLA are therefore valuable. Similar to the identification of optical counterparts for radio and

X-ray sources, we would like to estimate the probability that a given galaxy is associated with a DLA, given some observables (e.g. the impact parameter, the H I column density, etc). We will use two different treatments for this purpose: (i) a frequentist approach, used to test whether a candidate is an interloper, and (ii) a Bayesian estimator used to assign a probability that a candidate is actually associated with the DLA. Considered jointly, they can help to decide which galaxy (if any) is the DLA host, without the limitations imposed by color determinations or galaxy brightness.

We stress that we do not aim to provide a secure galaxy identification by this approach. Nevertheless, this technique is useful to pre-select the best candidates for spectroscopic follow-up. Also, when the present and future searches will yield a significant number of spectroscopically-confirmed galaxies, one can refine the statistical methods introduced here to select a *bona fide* DLA sample, useful to study the properties of the DLA population rather than those for individual detections.

4.6.1 Frequentist approach

The frequentist method is based on Poisson statistics applied to number counts of the surface density of galaxies. For each candidate with an impact parameter b and apparent magnitude m, we can compute the probability of detecting one interloper in the parameter space ($< b, < m$). Low values for this probability indicate that the candidate is unlikely to be an unrelated object, suggesting that it is likely to be the DLA host or the host of a second absorber at lower redshift along the sightline.

Given the surface number density of objects brighter than a fixed magnitude $\bar{n} = n(\leq m)$, the mean number of interlopers expected for $r \leq b$ is

$$\rho = \pi b^2 \bar{n} \tag{4.3}$$

and the probability to detect at least one galaxy is (Downes et al. 1986)

$$P_f = 1 - e^{-\rho} \, . \tag{4.4}$$

If $P_f \ll 1$, it is unlikely that the candidate corresponds to an interloper. However, as widely discussed in the literature (e.g. Downes et al. 1986, Sutherland & Saunders 1992), the probability of the candidate being the right identification does not follow immediately as $1 - P_f$. In fact, when multiple candidates lie within the search radius, the probability for each object is computed independently, leading to the ill-defined case in which the total probability for all candidates is not unity. The correct probability of a galaxy-absorber association comes from a Bayesian treatment (see next section).

For $\bar{n}$, we use galaxy number counts derived by Grazian et al. (2009) (their Table 1; see also Rafelski et al. 2009) from U-band imaging in a wide sky region (~ 0.4 sq. deg.), down to $U = 27.86$ AB mag.; this limit matches the depth of our survey. In Figure 4.5, we plot the dependence of P_f on the impact parameter, derived from Eq. (4.4) for different magnitude cuts. This analysis outlines how two competing effects play a role: depth and confusion. In fact, deep imaging is desirable to increase the chance of detecting absorber counterparts, but at the same time the number density of interlopers increases steeply, introducing significant confusion even at low impact parameters. Specifically, deep surveys, to u magnitudes fainter than ~ 27 mag., have a significant probability ($> 20\%$) of finding interlopers at impact parameters $b_{as} > 2''$. This result stresses the need for a quantitative treatment to identify DLAs: simply assuming the closest candidate to the quasar sightline to be the counterpart can lead to false detections. Note, however, that the figure also shows that this assumption is likely to be a good one for $b_{as} < 1''$ (corresponding to a physical distance of < 8.2 kpc at $z = 2.5$).

Table 4.3 lists the frequentist probability P_f for all candidates in the fields of J211444-

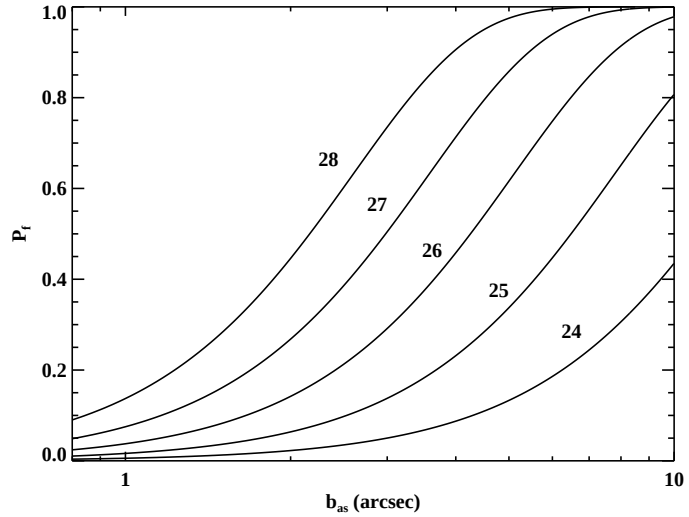

Figure 4.5: The probability P_f of detecting at least one interloper as a function of the impact parameter b_{as}, from Eq. (4.4). Different lines are for different cuts in u magnitude. In very deep surveys, there is a significant likelihood of detecting an interloper for $b_{as} > 2''$.

005533 and J073149+285449. In the latter field, candidate-A has a low probability of being an interloper ($P_f = 0.06$), making it the most likely candidate for the DLA host. Conversely, for J211444-005533, the minimum probability ($P_f = 0.08$) is for candidate-F, a bright galaxy at $z \sim 0.3$ (as measured in the SDSS and confirmed by our photometric redshift analysis), while the next lowest probability of being an interloper ($P_f = 0.19$, for candidate-A) is non-negligible. This case suggests that the frequentist analysis needs to be complemented with additional priors on the impact parameter, in the absence of spectroscopic information. In fact, as we will show in the next section, it is quite unlikely that luminous DLA galaxies lie at very large distances (~ 60 h_{72}^{-1} kpc for J211444-005533-F) from the quasar sightline.

4.6.2 Bayesian approach

Although independent of any priors, the frequentist approach does not yield a relative probability that an identification is correct, nor does it take into account the fact that several

candidates can be considered for a single DLA. Both of these can be achieved by a Bayesian treatment. The present section is organised as follows: after a review of the basic Bayesian formalism, we propose a method that can be applied to identify the host galaxies of generic absorption line systems (ALSs), such as Mg II absorbers, LLSs or DLAs. We then derive specific priors on the impact parameters of DLAs based on theory and indirect observational constraints. In the end, after testing this procedure on a sample of six spectroscopically-confirmed DLA hosts, we apply the method to compute probabilities for our targets in the fields of J211444-005533 and J073149+285449.

4.6.2.1 Formalism

Different approaches based on Bayes' theorem have been developed to identify the optical counterparts of X-ray or radio sources and, more recently, sub-mm sources. Due to the large number of works focused on this topic, varying terminology has been introduced over the past years. To make explicit our choice, we will review the fundamental concepts at the base of this method, mostly following Rutledge et al. (2000). Further, we also optimize the procedure for the issue addressed in this paper, namely the identification of the galaxy counterparts of high-z absorbers.

For a group of M candidates, the likelihood ratio LR is defined as the product of the normalised probability distribution functions (PDFs) of some properties x_{als} of the ALSs to those of random foreground[6] galaxies. The useful physical quantities are various observable parameters, including magnitude, impact parameter, H I column density, metal line equivalent widths, and kinematics. However, while the inclusion of many properties enables a narrower distribution of the likelihood ratio which restricts the number of false detections, this method

[6]Due to the design of our experiment, objects at redshifts higher than the LLS one cannot be detected in the proximity of the quasar. Therefore we define interlopers as foreground galaxies, even though this is not entirely appropriate for objects at redshifts between those of the DLA and the LLS.

is sensitive to the functional form adopted for x_{als}. To avoid subtle biases, it is hence better to restrict the number of priors to only well-known quantities. Here we consider a simple case in which only priors on the impact parameter $f(b)$ and magnitude distribution $q(m)$ are assumed.

Following Sutherland & Saunders (1992), we define LR as

$$LR = \frac{q(m)f(b)}{n(m)} \, . \tag{4.5}$$

LR is the ratio of the probability p of detecting a real counterpart at an impact parameter b and magnitude m

$$p = q(m)f(b)2\pi b \, \mathrm{d}b \, \mathrm{d}m \tag{4.6}$$

to the probability of detecting a random foreground object

$$p = n(m)2\pi b \, \mathrm{d}b \, \mathrm{d}m \, , \tag{4.7}$$

where $n(m)$ gives the distribution of galaxy number counts per unit area. This last quantity is not related to the nature of any particular ALS, and it can be derived empirically from deep imaging; as with the frequentist approach, we use the result of Grazian et al. (2009).

According to Bayes' theorem, the reliability R of a correct identification is

$$R_{als}(LR) = \frac{P(\mathrm{true}, LR)}{P(\mathrm{true}, LR) + P(\mathrm{false}, LR)} \, , \tag{4.8}$$

which is the ratio of the probability of true associations to the sum of true and false associations. R_{als} expresses the probability that a candidate with a given LR is the correct identification and not an unrelated foreground object. As pointed out by Sutherland & Saunders (1992), equation (4.8) does not account for the fact that multiple candidates can be considered for a

single absorber. In other words, a high value of R indicates that the considered candidate is an unusual source compared to the foreground galaxies, but frequently high reliability is assigned to more than one object. Eq. (4.8) provides no insight to solve this ambiguity.

To add this missing information, we introduce two other statistics. The first one is the probability $P_{no,id}$ that none of the M possible candidates is associated with the ALS:

$$P_{no,id} = \frac{\Pi_{j=1}^{M}(1 - R_j)}{S} \, . \tag{4.9}$$

The second is the probability $P_{als,i}$ that the i-th source is uniquely associated with the ALS:

$$P_{als,i} = \frac{R_i \Pi_{j \neq i}(1 - R_j)}{S} \, . \tag{4.10}$$

In the previous two equations, S is a normalization factor that ensures that $P_{no,id} + \sum_{i=1}^{M} P_{als,i} = 1$:

$$S = \sum_{i=1}^{M} R_i \Pi_{j \neq i}(1 - R_j) + \Pi_{j=1}^{M}(1 - R_j) \, . \tag{4.11}$$

In the end, Eq. (4.10) is the quantity that will be used to identify likely galaxy-absorber associations.

Before we apply this procedure to the case of DLAs, we highlight a possible problem that can affect the computation of LR with Eq. (4.5). For ALS studies, the form of the prior $q(m)$ has to be chosen carefully. Properties of ALSs in emission are currently poorly constrained and very little or nothing can be inferred about $q(m)$ from observations. Simulations can only partially help, especially because the star formation rate and stellar emission here are mostly computed based on semi-empirical prescriptions; this implies that any priors derived from simulations may not be reliable. Conversely, the use of the observed luminosity functions of high-redshift galaxies might imply a strong *a priori* constraint on the nature of

the ALS counterparts. Furthermore, although with significant noise, $q(m)$ can be obtained in a statistical sense by subtracting the magnitude distribution of galaxies in fields without DLAs from that in fields with known absorbers. This procedure requires a significant number of fields for convergence, but these observations are currently unavailable. Without a reliable estimate for $q(m)$, we suggest reducing the number of priors in the likelihood ratio, rather than adopting an inappropriate choice that might introduce uncontrolled biases. Note that one of the goals of our survey is to characterise the star formation properties of DLAs; incorrect information on the magnitude prior might have significant implications for the final result.

To remove $q(m)$ from the likelihood ratio, following Sutherland & Saunders (1992), we modify the definition of the likelihood ratio by marginalising Eq. (4.5) over m. We define LR_{als} as

$$LR_{als} = \frac{Q(m_l)f(b)}{M(m_l)}\,, \tag{4.12}$$

where

$$Q(m_l) = \int_{-\infty}^{m_l} q(m)\mathrm{d}m \tag{4.13}$$

and

$$M(m_l) = \int_{-\infty}^{m_l} n(m)\mathrm{d}m \tag{4.14}$$

are the priors $q(m)$ and $n(m)$, integrated up to the limiting magnitude m_l. Both Q and M are constants; the fact that $q(m)$ is unknown implies that the likelihood ratio has now an unspecified normalisation. Therefore, we adopt an operational definition of Eq. (4.8) as the probability of not obtaining $R_{als,i}$ randomly for the i-th candidate (Gilmour et al. 2007). The idea behind this procedure is to compute a distribution for LR using several sets of interlopers (N_{int}). High reliability is assigned to candidates whose LR exceeds typical values found among interlopers.

Formally, this is granted by

$$R_{als,i} = 1 - \frac{N(LR > LR_{als,i})}{N_{int}} \,, \tag{4.15}$$

where $N(LR > LR_{als,i})$ is the number of interlopers with a likelihood ratio that exceeds $LR_{als,i}$. N_{int} should be large enough to guarantee the convergence of $R_{als,i}$. Because the condition $LR \geq LR_{als,i}$ in Eq. (4.15) is satisfied modulo an arbitrary positive constant, the final reliability is independent of Q and N. Since the likelihood ratio distribution is computed directly from the imaging (see Sect. 4.6.2.3), this procedure offers the additional advantage of treating the limiting magnitudes m_l self consistently. The downside of this empirical approach is that we lose knowledge on $n(m)$, a well-defined quantity. However, we complement the Bayesian treatment with the frequentist approach, which includes the number count statistics.

4.6.2.2 Impact parameter modelling

Considering the specific case of DLAs, the only unspecified quantity at this point is the prior on the impact parameter $f(b)$. This can be derived from observations if a sample of spectroscopically-confirmed objects is available. Unfortunately, the hosts of only six DLAs at $z \gtrsim 2$ have so far been confirmed with spectroscopy, implying that it is not currently possible to use the observed impact parameters to constrain $f(b)$. We will hence indirectly derive a prior on b; the downside is that the final derived probabilities will carry additional uncertainty. In the future, an updated form of $f(b)$ derived directly from observations can provide a more reliable prior for statistical analysis.

Here we introduce and compare two different priors. The first one is based on the ΛCDM cosmology framework, in which galaxies assemble through a series of minor and major mergers. During this process, gas is thought to be distributed in clumps and filaments which

do not necessarily resemble low-redshift disks. In addition, gas in individual halos can cool to form a disk whose size follows the size evolution of the dark matter halo. For this reason, we refer to this prior as "evolutionary". Conversely, our second prior is based on the observational results of Prochaska & Wolfe (2009), who used a large DLA sample ($\sim$ 1000 DLAs) to find that the shape of the frequency distribution of projected H I column densities $f(N_{\rm H\,I}, X)$ does not evolve significantly with time at $z > 2$ and also matches the one at $z \sim 0$. This implies that the convolution of the projected H I surface density distribution in individual DLAs, their sizes and number density is preserved over $\sim$ 10 Gyr. A stronger interpretation presented by Prochaska & Wolfe (2009) is that $z \sim 3$ galaxies have H I disks whose distribution matches that of present-day spirals, a result that in turn suggests how H I in individual galaxies could be not especially sensitive to the underlying dark matter distribution. Therefore, we refer to this second prior as "non-evolutionary". Further investigations are required to confirm or disprove this hypothesis, but for now we note that the non-evolutionary prior is also useful to account for more extended H I than the one found inside simulated disks at high redshift.

We construct the non-evolutionary prior $f_{obs}(b)$ by simulating the quasar experiment, using sightlines through H I-21cm maps of local galaxies to reproduce the DLAs seen against background quasars (c.f. Zwaan et al. 2005a). For this purpose, we use H I-21cm maps from the THINGS survey (Walter et al. 2008) which includes 22 spirals and 12 Sm/dIrr galaxies at a resolution of $\sim 7''$. For each galaxy, we measure the local H I column density by averaging the signal within a resolution element in $\sim$ 200 random positions. We re-project the observed $N_{\rm HI}$ to a variety of inclinations, thus accounting for the fact that face-on disks are more likely to be selected in absorption than edge-on ones. For H I column densities above the DLA limit, we then compute $f_{obs}(b)$ by combining all the different sightlines for each galaxy. To reproduce a population of galaxies, we weight each object in the THINGS sample with the H I mass function (HIMF) Θ (Zwaan et al. 2005b) and the sky covering factor A. This is computed assuming that

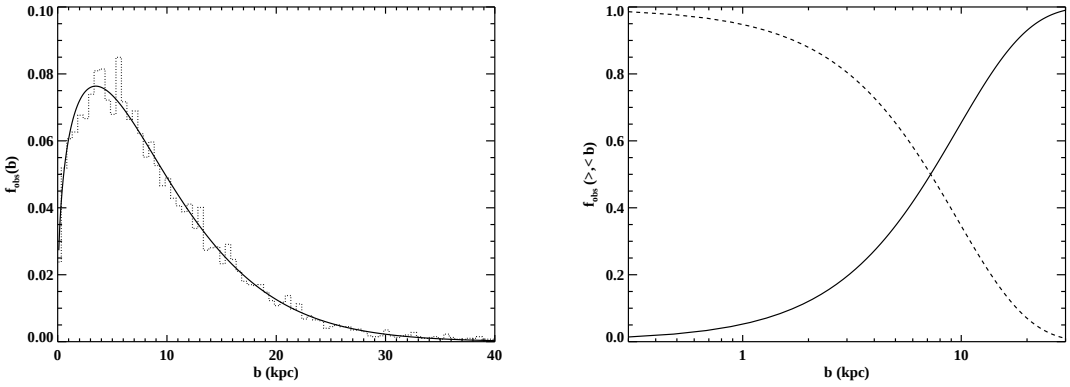

Figure 4.6: Left: Probability distribution function $f_{obs}(b)$ in one realisation of the quasar experiment (dashed histogram) and for the model OB0 (solid line). Right: Cumulative distributions $f_{obs}(> b)$ (dashed line) and $f_{obs}(< b)$ (solid line). According to this prior, $\sim 50\%$ of DLAs at $z = 3$ are expected to lie within $1''$ from the quasar, with the maximum probability located around $0.5''$.

the H I radius scales with the H I mass (Verheijen 2001). We also include a correction factor proportional to the number of galaxies (N_{gal}) in a given mass range (that defines N_{bin} bins) to compensate for the fact that dwarf galaxies are undersampled in the THINGS survey with respect to spirals. Combining all of these elements, we derive $f_{obs}(b)$ with

$$f_{obs}(b) = \frac{\sum_{i=1}^{N_{bin}} \frac{1}{N_i} \left(\sum_{k=1}^{N_{gal,i}} \Theta(M_k) A(M_k) f_k(b) \right)}{\sum_{i=1}^{N_{bin}} \frac{1}{N_i} \left(\sum_{k=1}^{N_{gal,i}} \Theta(M_k) A(M_k) \right)} \tag{4.16}$$

It is worth mentioning that high-redshift DLAs do not probe exclusively sight-lines similar to the ones through local disks as seen in 21cm. In fact, highly ionized species (e.g. N V) most likely associated with the halo are sometimes observed (Fox et al. 2009). However, the use of local 21cm maps seems an appropriate analogy to model H I-rich galaxies at high redshift.

In the left panel of Figure 4.6, we show $f_{obs}(b)$ (dashed histogram) from one realisation of the above equation (4.16). To model this distribution, we fit a function of the form

$$f(b) = A b^{\alpha} \exp(-B b^{\beta}) \, . \tag{4.17}$$

This analytic formula is designed to reproduce $f(b)$ for local galaxies: the power law accounts for the increasing probability of intersecting a disk at larger radii, while the exponential term accounts for the radial decay of the H I surface density profiles. The solid line shows the fit computed over 50 such experiments; the derived parameters and statistical uncertainties for this non-evolution model (OB0) are quoted in Table 4.4. It is reassuring that, although we are using a smaller sample, $f_{obs}(b)$ resembles qualitatively the distribution derived by Zwaan et al. (2005a). In the right panel of Figure 4.6, we plot the cumulative distributions $f_{obs}(> b)$ (dashed line) and $f_{obs}(< b)$ (solid line), obtained from the OB0 model. From this analysis, we infer that $\sim 50\%$ of DLAs at $z = 3$ are expected within $1''$ from the quasar with the maximum probability located around $0.5''$. Since DLAs are expected with low probability at impact parameters ≥ 40 h_{72}^{-1} kpc ($\sim 5''$ at $z = 3$), our search radius of $12''$ seems large enough to guarantee sufficient sky coverage during our candidate selection.

Turing our attention to the evolutionary prior, we derive $f_{sim}(b)$, using a cosmologically-weighted sample of DLAs drawn from the SPH simulation of Pontzen et al. (2009). This is similar to the simulation presented in Governato et al. (2007) and analysed in Pontzen et al. (2008), but at higher resolution. According to these authors, the impact parameter is defined as the projected distance to the minimum of the dark-matter halo potential. This is not an observable quantity, but it is reasonable to assume that high star formation occurs when the gas funnels towards the center of the halo, so that this definition of the impact parameter does not yield different results from our observational one. It is useful to note that by selecting individual halos, we are considering the gas distribution inside individual galaxies, with no distinction between the central galaxy and satellites (see Sect. 4.6.2.3).

The dashed histogram in the left panel of Figure 4.7 shows $f_{sim}(b)$ from a realisation of DLAs at $z = 3$ from the SPH simulations of Pontzen et al. (2009). We model $f_{sim}(b)$ (solid line) using the fitting formula in Eq. (4.17); although designed for local galaxies, this function

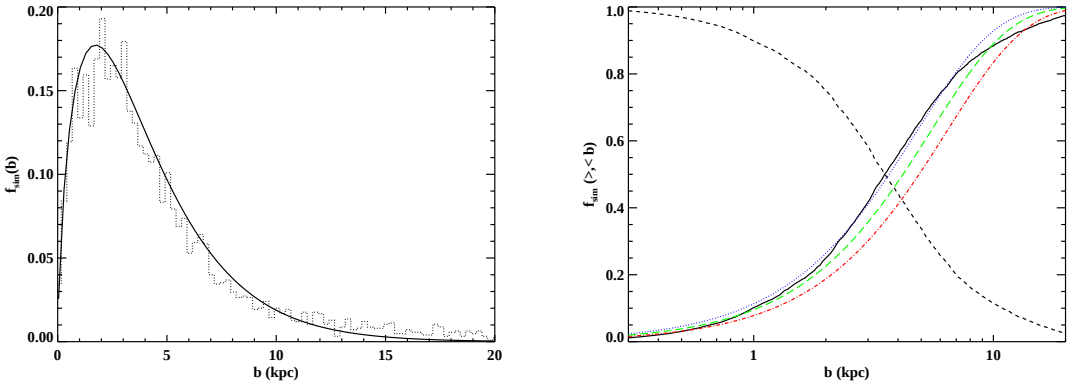

Figure 4.7: Left: The probability distribution function $f_{sim}(b)$ from an SPH simulation at $z = 3$ (dashed histogram) and for the fitted model SB3 (solid line). Right: Cumulative distributions $f_{sim}(> b)$ (dashed line) and $f_{sim}(< b)$ (solid line). Also shown, cumulative distributions derived with a toy model for $z = 3$ (OB3), $z = 2.5$ (OB2.5), and $z = 2$ (OB2) (blue dotted, green long-dashed, and red dash-dotted lines). According to this prior, $\sim 50\%$ of DLAs at $z = 3$ are expected to lie within $0.5''$ from the quasar with the maximum probability located around $0.3''$.

seems flexible enough to describe also the shape of $f_{sim}(b)$ in high-redshift ΛCDM simulations. The only discrepancy with the data arises at high b. The fitted parameters for this evolutionary model (SB3) are quoted in Table 4.4. In the right panel of Figure 4.7, we plot the cumulative distributions $f_{sim}(> b)$ (dashed line) and $f_{sim}(< b)$ (solid line), as obtained directly from the data. From these simulations, we deduce that a DLA can be found with $\sim 50\%$ probability within an impact parameter of $\sim 0.5''$ at $z = 3$. The maximum probability is located around $0.3''$, roughly a factor of 2 less than that predicted in a non-evolutionary model. These low values stress the advantage of our drop-out technique in observations limited by seeing. From a comparison of $f_{sim}(< b)$ with the rate of incidence derived by Nagamine et al. (2007), we note a similarity with their no-wind run, although our distribution exhibits a narrower tail.

The major difference between models OB0 and SB3 resides in the fact that gas follows the dark-matter potential more closely in simulations than in the non-evolving model, which assumes that the gas distribution does not change with redshift. This is reflected in a distribution for OB0 that is broader and peaks at higher impact parameters than the one for SB3. To

extrapolate the simulation results to $z < 3$, we scale the H I distribution observed at $z = 0$ following the size evolution of dark matter halos as a function of redshift. Starting with H I-21cm maps of local galaxies, we repeat the quasar experiment as for OB0, but this time accounting for a redshift dependence of b. In this toy model, we keep the observed surface density distribution constant, assuming that the total H I mass in the halo increases due to gas accretion onto the disk. In the literature, several scaling relations for the galaxy radius as a function of the redshift can be found, from both theoretical arguments and observations of high-z galaxies. In our model, we adopt $r(z) \propto H(z)^{-2/3} \propto (1+z)^{-1}$ (Bouwens et al. 2004) for $z > 1$. This is in agreement with Papovich et al. (2005) who show that the size distribution of galaxies at $z \lesssim 1$ is broadly consistent with that observed in the local Universe. This particular choice enables us to reproduce almost perfectly the SM3 model, starting from H I-21cm maps of $z = 0$ galaxies. This agreement is shown in the right panel of Figure 4.7, where the extrapolated model (OB3) is shown with a blue dotted line. The only significant discrepancy arises for $b > 7$ h_{72}^{-1} kpc. The figure also includes the cumulative distribution f_{sim} at $z = 2.5$ (OB2.5) and $z = 2$ (OB2) (green long-dashed and red dash-dotted lines, respectively). The parameters of the analytic expression (4.17) are listed in Table 4.4.

Table 4.4: Models for the impact parameter distribution from Eq. (4.17). OB0 is from H I maps of local galaxies and SB3 from SPH simulations at $z = 3$. OB2, OB2.5 and OB3 assume redshift evolution of local H I disks for $z = 3$, 2.5 and 2.

Type	A	α	B	β
OB0	0.064±0.001	0.37±0.02	0.057±0.009	1.29±0.05
SB3	0.234	0.68	0.37	1.05
OB2	0.112	0.37	0.10	1.29
OB2.5	0.138	0.37	0.12	1.29
OB3	0.166	0.37	0.14	1.29

Being able to match the simulations with an *ad hoc* $r(z)$ may not seem an interesting

result. However, other scaling relations (e.g. Ferguson et al. 2004) are equally able to reproduce a distribution at least consistent with the simulations. We speculate that there might be a more profound reason for this agreement: gas clumps re-assemble in growing dark matter halos without a drastic change in the radial distribution of the H I column density since $z = 3$ or even beyond. Further investigations on the gas distribution in SPH simulations are desirable to investigate this hypothesis. While accounting for disk evolution, we have assumed that the weighting procedure defined in Eq. (4.16) does not change as a function of redshift. Note that the impact parameter distribution is a normalised quantity, and is hence not much affected by any mass-independent variation in the number density of H I galaxies or in the covering factor A. Conversely, a change in the slope of the HIMF may alter the relative contribution of massive and dwarf galaxies, altering $f(b)$. No direct determinations of the HIMF as a function of redshift are currently available, and we hence keep the slope constant, consistent with the semi-analytic model of Obreschkow & Rawlings (2009) for $z < 3$ (see their figure 1).

4.6.2.3 Implementation, procedure test and discussion

Once the priors on the impact parameter are known, we test this Bayesian procedure using a sample of six spectroscopically-confirmed high-redshift DLAs (see Appendix 4.10.2). Although heterogeneous, this sample provides the only present observational test. To evaluate the reliability with Eq. (4.15), we need a realisation of LR for foreground objects. For this purpose, we compute LR for all the galaxies detected within $r_{search} < 10''$ of a random position in a field where no absorbers are known. We repeat this procedure for several random positions to guarantee the convergence of LR. Since we restrict to a searching area $r < 10''$, we implicitly impose the condition $LR_{dla,i} = 0$ for $b > 10''$. This well-defined boundary prevents probability from flowing towards high impact parameters. Note that, if r_{search} is allowed to increase to arbitrarily large radii, the number of interlopers at large impact parameters with small but

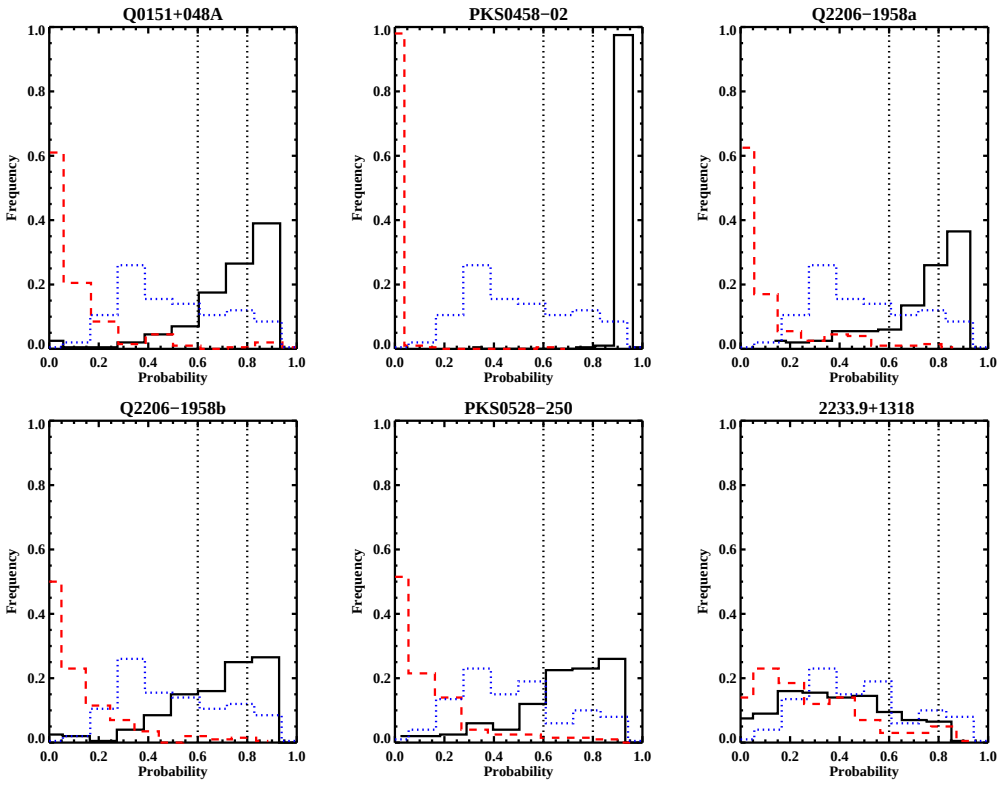

Figure 4.8: The results from 200 trials of the Bayesian procedure, with the probability assigned with the evolutionary prior to 6 known DLAs from the literature. The probability $P_{dla,i}$ assigned to the correct DLA is shown with a solid line, and that for the interloper with the highest reliability with a red dashed line. The blue dotted line indicates the highest probability assigned to foreground galaxies in a control test with no DLAs in the field. In all but one case (the SLLS towards 2233.9+1318), the Bayesian method assigns the highest probability to the correct galaxy-DLA association. The vertical dotted lines indicate $P_{dla,i} = 0.8$ and $P_{dla,i} = 0.6$. From this analysis, we infer that, in an ideal experiment, we expect to detect 60 *bona fide* counterparts out of 100 fields which host detectable DLAs when we assume the criterion $P_{dla,i} > 0.8$, while 15 interlopers will be incorrectly classified as DLAs.

non-zero LR will increase accordingly. Therefore, a non-zero reliability will be assigned also to DLA candidates with a low likelihood ratio, effectively decreasing $P_{dla,i}$ for the most likely candidates. This issue is bypassed by limiting the search radius to $r < 10''$.

After this, we extract all the sources detected within $10''$ from a random position in a field not hosting any known DLAs. Then, we add to this list of interlopers a known DLA at its measured impact parameter; finally, we compute $P_{dla,i}$ and $P_{no,id}$ for all of these candidates using both the evolutionary (SB3, OB2.5 and OB2) and the non-evolutionary (OB0) priors. We repeat this test 200 times for each confirmed DLA. To estimate the number of interlopers that are incorrectly identified as DLAs, we also run a control test in which only foreground sources are included. The results are in Figures 4.8 and 4.9 where we compare results for the evolutionary and non-evolutionary priors, respectively. For each known DLA, the probability $P_{dla,i}$ assigned to the correct galaxy counterpart is indicated by a solid line, while that assigned to the interloper with the highest reliability is shown with a red dashed line. Finally, we display the results of the control test in fields without DLAs: the blue dotted line represents the probability $P_{dla,i}$ assigned to the foreground galaxy with highest reliability when only interlopers have been detected.

Several pieces of useful information can be derived from the plotted distributions. First, looking at the six panels in Figures 4.8 and 4.9, it is evident that in all but one case our procedure assigns the highest probability to the correct candidate DLA host. We therefore conclude that the Bayesian method is successful in finding the right galaxy-absorber association. The only evident failure is for the target 2233.9+1318, an SLLS with $N_{\rm HI}$= 20.0 cm^{-2}. As shown by Zwaan et al. (2005a) (see also Figure 4.13), the impact parameter is a decreasing function of the H I column density; using a prior derived for absorbers with $N_{\rm HI}\geq 2 \times 10^{20}$ cm^{-2} may hence underestimate the quasar-galaxy separation for absorbers with lower H I column densities by more than a factor of 2. Second, it appears that the evolutionary and the non-evolutionary priors reproduce similar values of probability. In fact, the relevant feature that distinguishes the

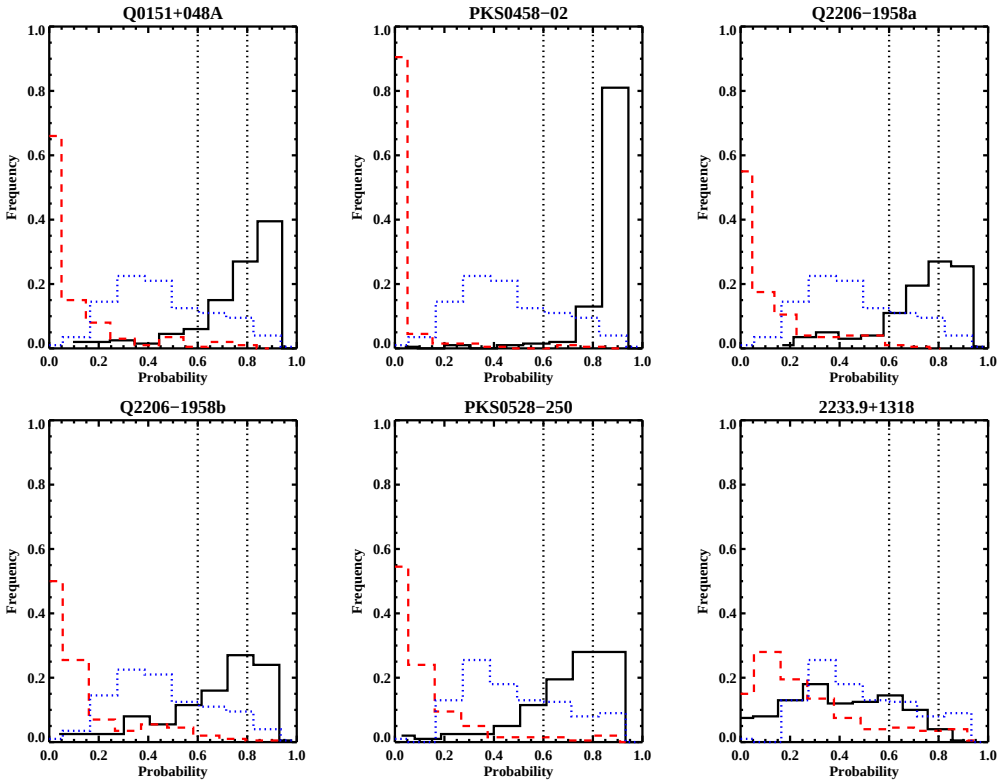

Figure 4.9: The same as Figure 4.8, but for the non-evolutionary prior.

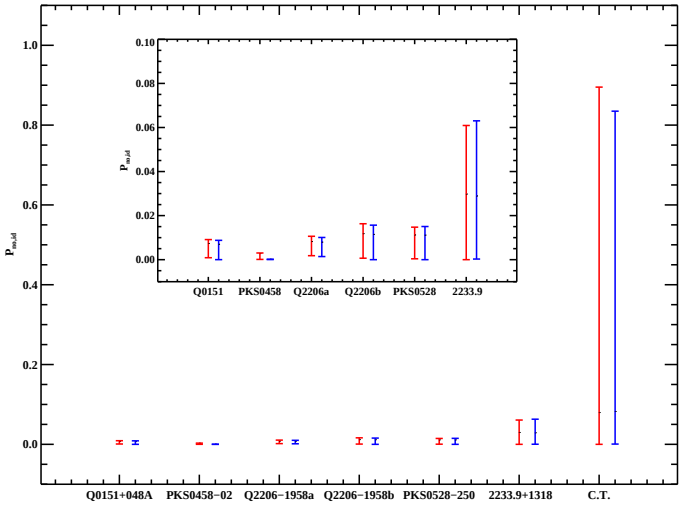

Figure 4.10: The probability of non detection ($P_{no,id}$) from 200 trials of the Bayesian procedure, on the left (in red) for the non-evolutionary prior and on the right (in blue) for the evolutionary prior. $P_{no,id}$ for the control test (C.T.; without known DLAs in the field) is highly dispersed, with a low mean value. In deep images, the total probability among the candidates is high due to the number density of interlopers and $P_{no,id}$ drops accordingly. The inset displays a zoom-in on the confirmed DLAs.

two models is the location of the maximum. From the definition of Eq. (4.15), similar values of $P_{dla,i}$ are expected with both priors when candidates lie on the tail of the $f(b)$ distribution.

It is also useful to note that the probability associated with the interlopers with highest R (red dashed line) peaks at low $P_{dla,i}$. This shows that Eq. (4.10) distinguishes between targets that share a high reliability. Finally, considering $P_{no,id}$ (Figure 4.10), we notice that this statistic is not particularly useful in deep imaging. In fact, although it behaves as expected for the fields with a known absorbers, for our control test in which only interlopers are included, $P_{no,id}$ exhibits a highly dispersed distribution with a mean value around 0.1. This is not exceedingly higher, although a factor of 10 larger, than the mean values $\sim 0.01 - 0.03$ derived from the other six experiments where DLAs are present. This is mostly due to the fact that very deep fields have a high number density of interlopers. When the number of candidates is large, even if no DLAs are detected in the field, $\sum P_{dla,i}$ can increase enough to make $P_{no,id}$ drop accordingly (since

$P_{no,id} + \sum_{i=1}^{M} P_{als,i} = 1$).

The fact that $P_{no,id}$ is not a useful indicator makes our analysis slightly more complicated. Our test shows that whenever a DLA is detectable in the field, the Bayesian procedure is able to correctly identify it (as shown by the comparison between solid black and dashed red histograms). However, if the DLA is too faint to be detected, interlopers may be incorrectly identified as the absorber, without any warnings from $P_{no,id}$. We quantify the number of spurious identifications by using our control test. For this purpose, we use the frequency with which high probability is assigned to interlopers in fields without DLAs. This provides an estimate of the contamination rate in our survey. Since the control test assumes that no DLAs are in the fields, this rate is somewhat overly-pessimistic. Finally, we note that our control test is not formally included in the Bayesian procedure, and the contamination rate we assume does not contribute to the probability $P_{no,id} + \sum_{i=1}^{M} P_{als,i} = 1$. Therefore, for a given probability limit P_{lim} on $P_{dla,i}$, the frequency with which a DLA is correctly identified (i.e., the number of trials for which $P_{dla,i} > P_{lim}$ in the solid histogram) and the frequency with which an interloper is incorrectly identified as the absorber (i.e., the number of trials for which $P_{dla,i} > P_{lim}$ in the dotted histogram) do not add up to one.

Nevertheless, these rates provide two extreme cases, useful to estimate the completeness and contamination in a *bona-fide* DLA sample derived with statistics. Our tests indicate that for galaxy-absorber associations with Bayesian probability $P_{dla,i} > 0.8$ the DLA galaxy is correctly identified $\sim 60\%$ of the time, whereas interlopers exceed $P_{dla,i} > 0.8$ only $\sim 15\%$ of the time. These rates have been computed excluding the SLLS towards quasar 2233.9+1318, not representative of the DLA population[7]. This means that such criteria should correctly identify 60 counterparts out of 100 fields with detectable DLAs. Conversely, in 100 fields that do not

[7]Due to the limited sample available, one might be concerned that these values are driven by the results for DLA PKS0458-02. However, even if we do not include this object, we find that still 55% of the galaxies are correctly identified, showing that the mean is not strongly dominated by this system.

show a DLA galaxy, these criteria would result in 15 interlopers being incorrectly classified as DLAs. If we weaken the probability limit down to $P_{dla,j} > 0.6$, our tests show that the 5 DLA fields have $P_{dla,i} > 0.6$ on average 85% of the time whereas interlopers exceed $P_{dla,i} > 0.6$ typically 35% of the time.

Turning to the discussion, we should emphasis that this statistical method is based on a set of assumptions that may not hold for all the sight-lines under consideration. We wish to discuss some of them in more details. First, the fact that we are considering a single galaxy-absorber association at a time can pose a limitation when a group of galaxies is located at the absorber redshift (e.g. towards Q2206−1958; Weatherley et al. 2005). In fact, our analysis will favor only one object and reject the other as interlopers. Conversely, clustering around the quasar (Hennawi & Prochaska 2007) and the QSO host galaxy itself do not affect our analysis since UV light from these galaxies is absorbed by the intervening LLS, as long as the systems are covered in projection by this absorber.

In addition, we cannot rule out with this statistical method that the detected objects are not associated with other intervening absorbers at $z < z_{lls}$. Indeed, towards QSO J211444-005533 we detect two Mg II systems at $z = 2.02$ and $z = 1.84$ and a C IV system at $z = 3.14$. Similarly, there are two Mg II systems at $z = 1.80$ and $z = 1.88$, towards J073149+285449. This source of confusion is partially alleviated by the fact that the priors on the impact parameters for Mg II peak at larger values. In fact, both observations and simulations (e.g. Kacprzak et al. 2009, Chen et al. 2010) show that Mg II are frequently (but not uniquely) found at $b > 20$ kpc.

As for the choice of the non-evolutionary prior, simulations show that massive halos can host multiple gas-rich satellites (e.g. Ceverino et al. 2010). In magnitude-limited surveys, only the brightest systems (central galaxies) will be detected, but also the satellites can give rise to an absorption line. Therefore, the most valuable quantity to set the priors may be the distance from each gas clump to the brightest star-forming center in the halo. In this configuration, a prior

will exhibit a more extended tail towards larger impact parameters than the one here presented. Future works will address this issue. For now, we caution that we will probe only those DLAs that originate within the brightest central star-forming centers. Indeed, the inclusion of larger impact parameters in this statistical procedure is not a trivial task: the number of foreground sources is a steeply increasing function of the distance from the quasar and the high degree of confusion is not optimal to identify this particular class of DLAs via statistics. Integral-field or multi-object spectroscopy down to faint magnitudes becomes essential.

Finally, we already mentioned a few times that a set of spectroscopically confirmed DLAs can be used to improve this statistical procedure. In order to establish how large a sample should be to determine $f(b)$, we extract randomly a subset of DLAs from the SPH simulation. While a large number of DLAs ($\sim$ 50-100 objects) is required to precisely reconstruct $f(b)$, a smaller sample ($\sim$ 20-30 objects) is sufficient to constrain the peak and the tail of the impact parameter distribution. Therefore, the present and other ongoing attempts to enlarge the sample of known DLAs may provide soon enough objects to improve this Bayesian procedure.

4.6.3 Results for J211444-005533 and J073149+285449

Before we compute $P_{dla,i}$ for our candidates, we remark on two points that have already been discussed. (1) Being a statistical analysis, this classification is subject to individual failures and carries all the assumptions and uncertainty related to the choice of the priors. (2) Due to the nature of our experiment, objects detected in the u band images at low impact parameters are at $z < z_{LLS}$. Therefore, the high-redshift LLS and the QSO host galaxy are not included in this analysis and they do not contribute to additional confusion. Additional confusion can arise from other absorbers (e.g. Mg II) in the line of sight.

Table 4.5: Bayesian statistics for DLA candidates in our fields. For J211444-005533 we use the priors OB0 (column 3,4) and SB3 (column 5,6). For J211444-005533 we adopt priors OB0 (column 9,10) and OB2.5 (column 11,12). *a* - R_{dla} expresses the probability that a candidate is the correct identification and not an unrelated foreground source. *b* - $P_{dla,i}$ is the probability that the i-th galaxy is uniquely associated with the DLA.

ID	b_p (kpc)	R^a_{dla}	$P^b_{dla,i}$	R_{dla}	$P_{dla,i}$	ID	b_p (kpc)	R_{dla}	$P_{dla,i}$	R_{dla}	$P_{dla,i}$
		J211444-005533						J073149+285449			
A	21.61	0.91	0.35	0.91	0.35	A	11.89	0.98	0.61	0.98	0.61
B	24.69	0.89	0.26	0.89	0.26	B	22.14	0.92	0.17	0.92	0.17
C	38.50	0.73	0.09	0.73	0.09	C	35.01	0.80	0.06	0.80	0.04
D	38.30	0.73	0.09	0.73	0.09	D	33.45	0.82	0.07	0.82	0.07
E	56.34	0.43	0.02	0.43	0.02	E	36.59	0.78	0.05	0.78	0.05
F	65.43	0.23	0.01	0.23	0.01	F	48.86	0.60	0.02	0.61	0.02
G	59.85	0.35	0.02	0.35	0.02	G	80.76	0.00	0.00	0.00	0.00
H	77.40	0.00	0.00	0.00	0.00	H	97.92	0.00	0.00	0.00	0.00
I	72.52	0.06	0.00	0.06	0.00	I	95.55	0.00	0.00	0.00	0.00
L	63.46	0.27	0.01	0.28	0.01	L	66.30	0.28	0.01	0.28	0.01
M	63.07	0.28	0.01	0.29	0.01	M	76.56	0.05	0.00	0.05	0.00
N	69.07	0.14	0.01	0.15	0.01	N	97.94	0.00	0.00	0.00	0.00
O	38.64	0.73	0.09	0.73	0.09	-	-	-	-	-	-
P	90.87	0.00	0.00	0.00	0.00	-	-	-	-	-	-
Q	89.79	0.00	0.00	0.00	0.00	-	-	-	-	-	-
R	90.78	0.00	0.00	0.00	0.00	-	-	-	-	-	-
S	97.75	0.00	0.00	0.00	0.00	-	-	-	-	-	-
T	91.73	0.00	0.00	0.00	0.00	-	-	-	-	-	-

Bearing these caveats in mind, but encouraged by the positive results from our tests, we apply the above statistical procedure to the galaxies detected in the fields of J211444-005533 and J073149+285449. Reliabilities and probabilities of galaxy-absorber association are listed in Table 4.5. For the $z \sim 2.919$ DLA towards J211444-005533, we use the templates OB0 and SB3, while for the $z \sim 2.686$ DLA towards J073149+285449, we adopt OB0 and OB2.5. For the DLA towards J211444-005533, our statistics indicate that none of the detected targets has a probability greater than 35% of being associated with the DLA. Conversely, in the case of J073149+285449, there is a probability of $\sim 60\%$ that object A is associated with the DLA. Adding the fact that the probability of being an interloper is less than 10% from the frequentist analysis, we consider J073149+285449-A an excellent candidate for the DLA host galaxy. We are presently trying to confirm the association in J073149+285449 through spectroscopy in the Lyα line and UV continuum.

4.7 SFR calibration

The metal lines observed in DLAs support the idea that star formation activity has occurred at least previously in these objects, enriching the surrounding gas (e.g. Wolfe et al. 2003). A key issue in DLA studies is the star formation rate in these objects and its distribution across the ISM of the host galaxy.

At $z = 3$, the u-band filter covers the rest frame wavelengths 740 Å $\lesssim \lambda \lesssim$ 1000 Å, where a galaxy's emission is expected to be dominated by massive ($M > 10 M_{\odot}$) and short-lived ($t_{life} < 2 \times 10^7$ yr) stars. To recover the emitted rest-frame UV flux $F_{\nu,e}$, we apply a simple K-correction to the observed flux $F_{\nu,o}$ under the assumption that the SED is not a sensitive

function of wavelength in the FUV region:

$$F_{\nu,o} = \frac{(1+z_e)L_{\nu,e}}{4\pi d_L^2} = (1+z_e)F_{\nu,e}\,, \tag{4.18}$$

where d_L is the luminosity distance to z_e. We also correct for absorption by the IGM, using an updated calculation of the effective opacity τ_{eff} computed from a recent determination of $f(N_{\rm H\,I}, X)$ (Prochaska et al. 2010) over a large interval of H I column densities ($10^{12} - 10^{22.5}$ cm^{-2}) at $z \sim 3.7$.

We compute the transmission T_{igm} to FUV photons considering the first 35 lines in the Lyman series as:

$$T_{igm}(\nu) = \exp[-\tau_{eff}(\nu)]\,, \tag{4.19}$$

where the effective opacity τ_{eff} is defined by

$$\tau_{eff} = \sum_{\nu_i} \int_0^{z_e} \int f(N_{\rm H\,I}, z)(1 - e^{-\tau_{c,i}}) {\rm d}N_{\rm HI} {\rm d}z\,, \tag{4.20}$$

with $\tau_{c,i}$ the optical depth of an individual cloud at the frequency $\nu = \nu_e(1+z)$ computed for the i-th element of the Lyman series with frequency ν_i. To relate $f(N_{\rm H\,I}, X)$ derived at $z_0 = 3.7$ by Prochaska et al. (2010) to $f(N_{\rm H\,I}, z)$ at an arbitrary redshift, we assume

$$f(N_{\rm H\,I}, z) = f(N_{\rm H\,I}, X)\frac{{\rm d}X}{{\rm d}z}\left(\frac{1+z}{1+z_0}\right)^{\gamma}\,, \tag{4.21}$$

where

$$\frac{{\rm d}X}{{\rm d}z} = \frac{H_0(1+z)^2}{H(z)}. \tag{4.22}$$

In Eq. (4.21), we model the redshift evolution in the interval $2 < z < 4$ with a density dependent power law index. Specifically, we assume $\gamma = 2.47$ for $N_{\rm HI} < 10^{17}$ cm^{-2} (Kim et al. 2002),

$\gamma = 2.78$ for $10^{17} < N_{\rm HI} < 10^{19}$ cm^{-2} (Prochaska et al. 2010), $\gamma = 1.78$ for $10^{19} < N_{\rm HI} < 2 \times 10^{20}$ cm^{-2} (Rao et al. 2006, O'Meara et al. 2007), and $\gamma = 1.27$ for $N_{\rm HI} > 2 \times 10^{20}$ cm^{-2} (Rao et al. 2006). A plot of the IGM transmission at redshifts 2, 3 and 4 is presented in the top panel of Figure 4.11 (solid lines), together with the u (blue dashed line), V (green dashed line) and R (red dashed line) LRIS filter transmission curves. Comparing the results of our calculation with those from a similar analysis by Madau (1995) (dotted line) at $z = 3$, we find that the major discrepancy arises for high order lines in the Lyman series. In fact, the main contribution to the opacity at these wavelengths comes mostly from optically-thick absorbers, which are more numerous in the Madau (1995) calculation than the estimate of Prochaska et al. A slight offset is also visible in the Lyman-α line, where our opacity is 2% higher than that computed by Madau (1995). Despite these differences, the two calculations for the transmission through a broad-band filter agree to within a few percent in the interval $2 \lesssim z \lesssim 4$. We derive the IGM correction C_{IGM} to an observed u-band flux by integrating the product of the effective opacity and the u-band transmission curve $g_u(\lambda)$

$$C_{IGM} = \int e^{-\tau_{eff}(\lambda)} g_u(\lambda) {\rm d}\lambda \, . \tag{4.23}$$

The final values, in magnitudes, are presented in the lower panel of Figure 4.11 as a function of redshift: the solid, dotted, and dash-dotted lines are for the u-, V- and R-bands, respectively.

Finally, we convert the UV luminosity L_ν (erg s^{-1} Hz^{-1}) into an SFR (M$_\odot$ yr^{-1}) using the calibration at 1500Å from Madau et al. (1998), divided by 1.58 to account for a Chabrier IMF (Salim et al. 2007):

$$SFR = 7.91 \times 10^{-29} L_\nu \, . \tag{4.24}$$

There are some caveats to this determination of the absolute SFR. First, it is not obvious

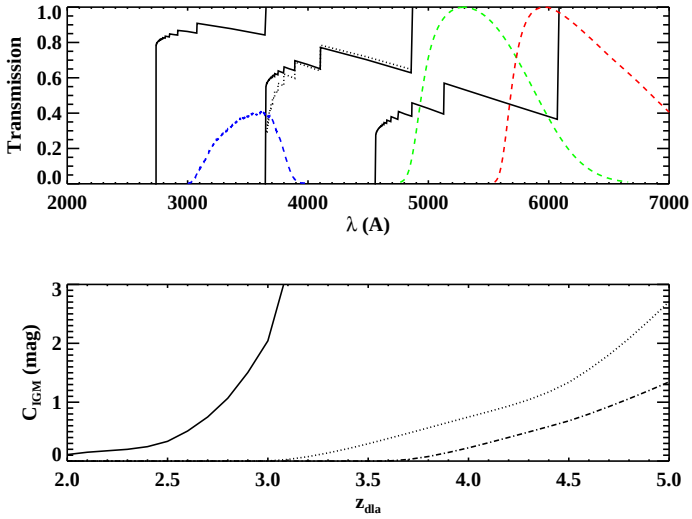

Figure 4.11: Top panel: The IGM transmission at $z = 2$, 3 and 4 (solid lines), together with the u (blue dashed line), V (green dashed line) and R (red dashed line) LRIS filter transmission curves. Superimposed is a comparison with the result of a similar analysis by Madau (1995) (dotted line) for $z = 3$. The major discrepancy arises for higher-order lines in the Lyman series due to the different numbers of optically-thick systems included in the two calculations. Bottom panel: The final IGM correction in magnitude as a function of the DLA redshift for the u- (solid line), V- (dotted line) and R-band (dash-dotted line) filters.

whether the conversion of Madau et al. (1998) is applicable at wavelengths lower than 1500Å. In fact, the presence of molecular gas in the ISM can significantly increase the opacity of the UV photons in the Lyman-Werner band, resulting in an underestimated SFR if the star forming regions of DLAs are rich in molecules. Unfortunately, it is very difficult to properly account for this effect. In addition, the SED is most likely frequency-dependent, especially towards harder UV frequencies. Comparing fluxes at 1000 Å and 1500 Å with the SED templates of Grazian et al. (2006), we find that $F_{1000}/F_{1500} \sim 1$ within a factor of 2 in scatter. *Ad hoc* calibrations can be computed, as done for example by Christensen et al. (2009). However, the same order of uncertainty is associated with different choices for the template age at a fixed metallicity. As a consequence, we infer that the absolute value of the SFR is uncertain at the level of a factor of 2. This is without considering additional complications due to dust extinction. Contamination from Lyα emission is not an issue for our two DLAs, as the Lyα line by design does not lie within the u band.

Applying our calibration, we derive for DLA J073149+285449-A an unobscured SFR of $(5.4 \pm 0.5 \pm 2.7)$ h_{72}^{-2} $M_\odot$ yr^{-1}, corrected by a factor of 1.9 due to IGM absorption. Here, the first uncertainty refers only to the error in the flux measurement, while the second one refers to a 50% uncertainty on the star formation calibration, combined with a 10% error from the IGM correction. To estimate the total star formation rate corrected for dust, one can include a factor of ~ 2.3, as suggested by X-ray measurements (Reddy & Steidel 2004) for galaxies with $SFR < 20$ h_{72}^{-2} $M_\odot$ yr^{-1}. For DLA J211444-005533, we derive an upper limit to the unobscured SFR of 1.4 h_{72}^{-2} $M_\odot$ yr^{-1}, computed at 3 σ C.L. in a $1''$ aperture, and corrected by a factor 4.4 for IGM absorption.

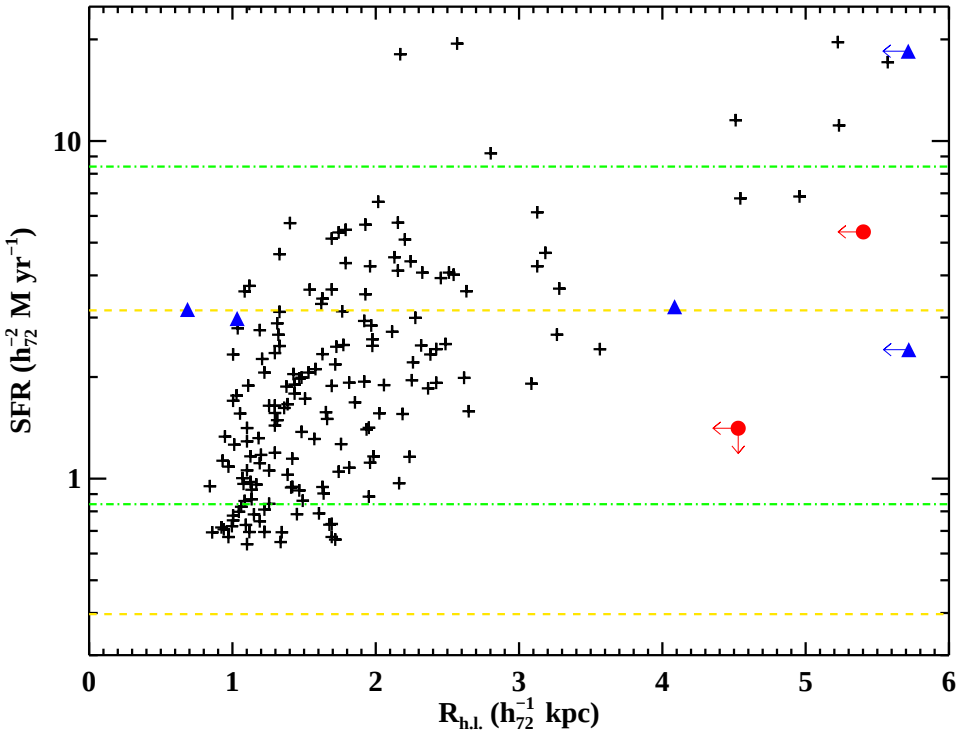

Figure 4.12: A comparison of SFRs and sizes between LBGs and DLAs. Crosses are for a sample of LBGs from Bouwens et al. (2004) while the blue triangles represent previously known DLA galaxies; the red circles are for DLA J073149+285449-A and the 3σ upper limit for the DLA towards J211444-005533. Also overplotted (in green dash-dotted lines) are a typical L_* and $1/10L_*$ LBG from Reddy & Steidel (2009) and the expected SFR in DLAs as inferred from [C II*] by Wolfe et al. (2003) (yellow dashed lines). Upper limits on the sizes are conservative estimates using the seeing FWHM for ground-based observations.

4.8 Discussion

4.8.1 Star formation rate in DLAs and LBGs

One of the outstanding questions in DLA studies is whether DLAs arise from the extended hydrogen reservoirs surrounding LBGs (e.g. Møller & Warren 1998). This idea is in agreement with the finding that Lyα emission is more spatially extended than the UV emission, suggesting that photons from newly formed stars are resonantly scattered before they can escape from the wings of the Lyα line (Rauch et al. 2008). It is also consistent with the finding that DLAs have too low an *in situ* star formation rate to justify the chemical enrichment and the cooling rate inferred from [C II*] absorption lines[8] (Wolfe et al. 2003). The possible link between DLAs and LBGs has been the subject of several studies, comparing the sizes, morphology and luminosities of the two classes of objects (e.g. Fynbo et al. 1999, Møller et al. 2002). It should be emphasized, however, that only one DLA, the $z \sim 1.92$ system towards Q2206−1958, has so far been directly shown to be associated with an LBG.

In Figure 4.12, we present an updated comparison of SFRs and sizes between LBGs and DLAs, using the SFRs derived for a sample of LBGs at $z \sim 3$ by Bouwens et al. (private communication; see also Bouwens et al. 2004). The LBGs are shown with crosses, while the blue triangles represent DLAs with spectroscopically-confirmed hosts (see Appendix 4.10.2). The red circles refer to candidate DLA J073149+285449-A and the 3σ upper limit for DLA J211444-005533, from this work. We also overplot with green dash-dotted lines a typical L_* and $1/10L_*$ LBG from Reddy & Steidel (2009) ($M^*_{1700\text{Å}} = -20.97$ at $z \sim 3$). Finally, the dashed lines indicate the expected range of SFRs in DLAs, as inferred from the [C II*] model of Wolfe et al. (2003). The lower value is for a cold neutral medium (CNM), while the higher one is for a warm

[8] It should be noted that the [C II*] model of Wolfe et al. (2003) is inconsistent with the H I temperature distribution in DLAs, as derived from H I-21cm absorption studies. See Kanekar & Chengalur (2003), Kanekar et al. (2009).

neutral medium (WNM); in both cases we assume a disk size of 100 kpc^2 to convert the SFR surface density into an integrated value. Regarding the DLA sizes, we assume the seeing FWHM as an upper limit on the half-light radius for our determinations; for previously-known DLAs, we quote half-light radii from Møller et al. (2002) for 3 galaxies, while we assume an FWHM of $0.8''$ as a conservative upper limit for the remaining cases. All of the quantities presented here have been rescaled to match our SF calibration, IGM correction and cosmology. This allows for a relative comparison which is not affected by the systematic uncertainty on the absolute value for the SFR.

Interestingly enough, J073149+285449-A lies in the same interval of SFR observed for the earlier DLAs. Although the statistics are still limited, as noted in previous studies, the detected DLAs appear to be consistent with the SFR distribution of LBGs, at least for the redshift interval considered. The only exception is the bright DLA towards Q0151+048A, whose gas is known to be photo-ionized by a nearby QSO (Fynbo et al. 1999). Due to the small sample and the many upper limits for sizes, we refrain from additional discussion here. We only comment on the fact that DLAs, being H I selected galaxies, are expected to span a wide range in UV luminosity (e.g. Pontzen et al. 2008). However, the optical follow-up of these objects imposes an additional selection bias since only the most luminous DLAs can be observed. We remark that non-detection from our imaging can only be attributed to the sensitivity limits and hence will directly constrain the DLA luminosity function.

Finally, considering the SFR surface density expected from the [C II*] cooling rate in DLAs, we note that a disk of 100 kpc^2 (lower yellow dashed lines in Figure 4.12) significantly underestimates the observed SFR, for the CNM model of Wolfe et al. (2003). For such a scenario, DLAs should typically arise in galaxies with star-forming regions extended over more than ~ 600 kpc^2, similar to present day disks. If this is the case, a significant number of DLAs may be at low-surface brightness. Conversely, if DLAs originate exclusively from a WNM, model

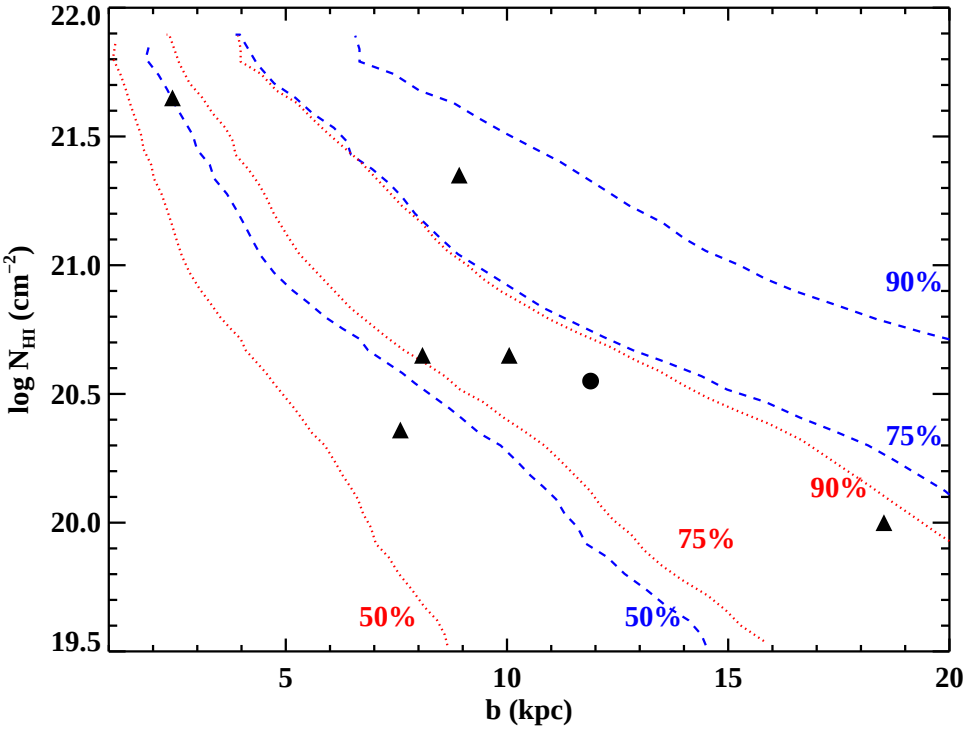

Figure 4.13: The H I column density as a function of the impact parameter for all confirmed DLA host galaxies (triangles) and for our candidate towards J073149+285449 (circle). Also overplotted with blue dashed lines are (from left to right) the 50^{th}, 75^{th}, and 90^{th} percentiles of the impact parameter as a function of the H I column density in local disk galaxies from Zwaan et al. (2005a). The same distribution scaled to $z = 2.35$ according to a simple model is shown with red dotted lines. Provided that previous observations were biased against small impact parameters, the H I in high-z DLAs seem more extended than what expected from single disks from simulations.

and observations agree for more compact disks, suggesting a typical DLA size of ~ 80 kpc^2 (close to the higher yellow dashed line in Figure 4.12). We hope to improve our knowledge of the DLA size through our upcoming HST (PI: O'Meara, ID 11595) observations.

4.8.2 The HI distribution at $z > 2$

One of the most valuable results we hope to achieve is to directly trace how neutral gas is distributed around star-forming galaxies at $z > 2$. This will address the fundamental

question of the origin of DLAs and ultimately provide important constraints for models of galaxy formation. Whereas next generation arrays such as the SKA will allow direct imaging of individual (albeit large) galaxies at $z \sim 2$ in the H I 21cm line, our analysis should, given a large enough DLA sample, constrain the radial distribution of H I in DLAs as a function of the projected distance to the star-forming center.

Evidence for an anti-correlation between the impact parameter and the H I column density has been obtained by previous studies aimed to identify DLA host galaxies (Møller & Warren 1998, Christensen et al. 2007). Similar results emerge from absorption line studies in QSO pairs or multiple images of lensed quasars (Monier et al. 2009). Once again, this is consistent with a model in which DLAs arise in gas located around star-forming galaxies. In Figure 4.13, we show the $b/N_{\rm HI}$ distribution for all DLAs whose impact parameters have been measured (triangles), along with our candidate DLA J073149+285449-A (circle). Our system lies within the population of confirmed DLA hosts, supporting previous claims for such an anti-correlation.

Figure 4.13 also overlays the above DLA data on the conditional probability of the impact parameter as a function of the H I column density, derived for local disk galaxies by Zwaan et al. (2005a; see their figure 15). After inverting $b \equiv b(N_{\rm HI})$, the 50%, 75% and 90% percentiles of this distribution are plotted with blue dashed lines in the figure. The probability is found to increase with increasing impact parameters, as seen tentatively in the high-z DLA sample. We also show (red dotted lines) the values expected after scaling the impact parameter (using the toy model of section 4.6.2.2) to $z = 2.35$, i.e. the median redshift of the observed DLAs. Compared with the local H I 21cm data, we find that almost all the high-z DLAs lie within 75% of the H I distribution seen in present-day galaxies. Conversely, DLAs are seen to lie outside the 75th percentile of the expected H I distribution for models in which the gas at high redshifts follows the dark matter halo potential. Interestingly, this result is consistent with

the gas distribution of lower redshift ($1 \lesssim z \lesssim 2$) absorbers (Monier et al. 2009). We emphasize that some of the previous observations are likely to have been biased against DLAs at small impact parameters; further, our toy model is based on very simplistic assumptions. Given these caveats, we infer that gas in high-z DLA galaxies is distributed on larger scales than those inferred from the simulated disks. This is suggestive that the null hypothesis that gas in high-z DLA galaxies has not changed its distribution from $z = 3$ to the present epoch is consistent with the present data, in agreement with the absence of evolution in $f(N_{\rm H\,I}, X)$, as pointed out by Prochaska & Wolfe (2009). While this agrees with the notion that DLAs are large disks at high redshifts, a model in which gas-rich satellites surround a central star-forming galaxy may explain the observed distribution equally well. In fact, even if gas is distributed in small disks inside individual halos, multiple satellites around a central galaxy could be responsible for an overall extended H I distribution.

4.9 Summary and future prospects

We have presented a new imaging programme, aimed to increase the sample of DLA host galaxies known between $z = 2$ and $z = 3$, based on an updated version of the dropout technique of O'Meara et al. (2006). We imaged QSO sightlines containing two optically-thick absorption line systems, using the Lyman limit of the higher-redshift absorber to block the quasar light at wavelengths shortward of the limit. This allows the rest-frame UV emission from the lower-redshift DLA to be imaged at wavelengths between the Lyman limits of the two absorbers, without any contamination from the QSO light. Using this criterion, we have selected a sample of 40 targets for HST and ground-based observations, which are currently being carried out. Once the locations of the candidate host galaxies have been determined from imaging, follow-up long-slit spectroscopy in either the Lyα or the Hα lines will allow a

measurement of the redshift of the DLA host. IFU spectroscopy will also reveal properties of the ISM (e.g. the metallicity or SFR via emission lines) from regions that are not probed by the QSO absorption lines. A significant advantage of the DLAs targeted in our survey is that, even for spectroscopy, the QSO contamination disappears blueward of the Lyman limit of the higher-redshift absorber. Any continuum detected in the spectrum comes only from foreground sources which are in the slit, increasing the probability of confirming the redshift of the galaxies associated with the low-z DLA.

In this paper, we have presented preliminary results from the application of this technique to sightlines towards two QSOs, J211444-005533 and J073149+285449, each hosting two high column density absorbers. Our Keck-LRIS u-band imaging of these fields achieved a depth of ~ 29 mag. at 1σ, resulting in the detection of a number of candidates for the host galaxy in each case. Follow-up spectroscopic studies are ongoing to confirm the galaxy-absorber associations for these and other candidate hosts identified in the programme.

To pre-select galaxies for spectroscopy or for those cases (e.g. very faint galaxies) where spectroscopic confirmation will be expensive, we have also proposed a statistical approach based on both number count statistics and a Bayesian treatment. This procedure is based on several simple assumptions and it does not aim to correctly identify each DLA, but provides a means to build a statistical sample of DLA galaxies that is representative of the entire population. We provide a general Bayesian identification procedure that can be applied to identify the galaxy counterparts of different types of absorption line systems, including DLAs. Due to the scarce present information on the impact parameter distribution in DLAs, we derived two different priors on this distribution, based on SPH simulations from Pontzen et al. (2009) and H I-21cm observations of local galaxies from THINGS (Walter et al. 2008). The second prior is computed under the assumption that the H I column density distribution does not evolve significantly with redshift (Prochaska & Wolfe 2009), while the first is more consistent with a hierarchical

picture of galaxy assembly. An observational determination of this prior will be possible once a larger sample (> 20 objects) of spectroscopically-confirmed DLA will be available, from the present and other ongoing surveys.

We have tested the proposed statistical approach on a sample of five DLAs and one SLLS whose host galaxies have been spectroscopically confirmed. For all DLAs, the procedure correctly identified the galaxy giving rise to the absorber; conversely, we could not identify the galaxy responsible for the sub-DLA (with $N_{\rm HI}= 10^{20}$ cm^{-2}), perhaps because our prior has been calibrated for higher H I column densities. Our test suggests that with this statistical method we can select a sample of *bona fide* DLAs complete to 60%-85% (depending on the required confidence level) with a contamination from interlopers around 15%-35%.

We then applied the proposed identification procedure to the candidate hosts detected in Keck-LRIS images of the fields of J211444-005533 and J073149+285449. For J211444-005533, no galaxies is found to be associated with the DLA at high significance, while for J073149+285449 we found a good candidate with 60% probability of being the DLA and less then 10% probability of being an interloper. This system is at a projected distance from the quasar $b = 1.54''$, corresponding to an impact parameter of 11.89 h_{72}^{-1} kpc at $z = 2.686$. Spectroscopic confirmation of this candidate is now being carried out. Converting the rest-frame UV emission into a SFR, we measure for candidate DLA J073149+285449-A an unobscured SFR of $5.4 \pm 0.5 \pm 2.7$ h_{72}^{-2} $M_\odot$ yr^{-1}, where the first uncertainty refers only to the error in the flux measurement, while the second one refers to a 50% uncertainty on the star formation calibration, combined with a 10% error from the IGM correction. Conversely, we place the 3σ upper limit of 1.4 h_{72}^{-2} $M_\odot$ yr^{-1} on the SFR of the DLA towards J211444-005533.

The SFR properties of our candidate and DLAs with identified hosts from the literature appear consistent with those of two independent samples of LBGs at similar redshifts (Bouwens et al. 2004, Reddy & Steidel 2009), supporting earlier suggestions that the brightest DLA galaxies

and LBGs might be overlapping galaxy populations. The impact parameter b measured for the new candidate DLA J073149+285449-A is also consistent with an anti-correlation between impact parameter and H I column density $N_{\rm HI}$, as suggested by earlier studies of DLAs and local galaxies. Comparing the b-$N_{\rm HI}$ distribution with the conditional probability of the impact parameter as a function of the H I column density derived in local disk galaxies, we find that most DLAs lie within the 75^{th} percentile. This is consistent with the an absence of redshift evolution in $f(N_{\rm H\,I}, X)$ at $z > 2$, consistent with the distribution at $z = 0$ (Prochaska & Wolfe 2009), but it could be equally well explained with systems composed by a central star-forming galaxy and multiple gas-rich satellites, in concordance with cold dark matter simulations.

With the direct images of high-z DLA host galaxies that will be available from our HST/Keck survey we aim to enlarge the current sample to answer some fundamental questions on the absorbers. What is the typical star formation rate in DLAs at $z \sim 2-3$? How are gas and stars distributed in the absorbers at these redshifts? Are rotationally-supported disks already in place at $z \sim 3$? Or are DLAs associated with merging gas clumps? How do the metallicity, the dust-to-gas ratio, and the H I column density depend on the impact parameter? Is there an SFR/metallicity or a mass/metallicity relation in DLAs? Over the next few years, it should hence be possible to obtain a comprehensive picture of the properties of high redshift DLAs, providing new insights into their role in the broad picture of galaxy formation and evolution.

4.10 Appendix

4.10.1 Photometric quantities

Quantities derived from the photometry of galaxies detected in the fields J211444-005533 and J073149+285449 are listed in Table 4.6. Relative separations from the quasar are in columns 2 and 3, in arcsec. Total magnitudes, uncertainties and signal-to-noise ratios for the

u band are in columns 4, 5, and 6, respectively. The remaining columns list the magnitudes computed in circular apertures in the R, V, and I bands. Additional details are provided in Section 4.4.2. All the listed magnitudes have been corrected for Galactic extinction.

Table 4.6: Table of photometric quantities. Asterisks indicate $3 < S/N < 5$. Magnitudes with $S/N < 3$ (including non-detections in a given band) are not listed in the table. The listed values have been corrected for Galactic extinction. See Appendix 4.10.1 for additional details on each entry.

ID	Δ ($''$)		u tot. (mag.)	σ_u (mag.)	S/N$_u$	u ap. (mag.)	$\sigma_{u,c}$ (mag.)	R ap. (mag.)	$\sigma_{R,c}$ (mag.)	V ap. (mag.)	$\sigma_{V,c}$ (mag.)	I ap. (mag.)	$\sigma_{I,c}$ (mag.)
J211444-005533													
A	2.54 E	1.32 N	25.53	0.08	15.8	25.62	0.11	24.31	0.10	24.49	0.10	24.25	0.11
B	3.19 W	0.73 S	26.98	0.15	7.5	-	-	-	-	-	-	-	-
C	1.45 E	4.88 N	25.58	0.09	14.3	25.84	0.08	24.64	0.06	24.65	0.05	24.69	0.09
D	0.54 E	5.05 S	27.78*	0.24	4.6	-	-	-	-	-	-	-	-
E	6.35 W	3.92 N	26.95	0.15	7.8	-	-	-	-	-	-	-	-
F	8.67 W	0.07 N	22.76	0.06	90.5	23.19	0.06	20.22	0.05	20.78	0.04	19.93	0.05
G	6.77 W	4.12 S	25.59	0.08	18.7	25.71	0.07	25.71	0.11	25.75	0.10	26.10	0.22
H	0.84 W	10.22 S	24.99	0.07	24.6	25.40	0.06	24.99	0.06	25.09	0.06	24.97	0.09
I	3.81 E	8.82 S	24.44	0.06	41.3	24.75	0.06	24.63	0.06	24.74	0.05	24.63	0.08
L	7.88 E	2.93 S	27.93*	0.30	3.7	-	-	-	-	-	-	-	-
M	8.35 E	0.374 N	25.33	0.09	14.9	26.10	0.08	25.98	0.13	25.99	0.11	26.02	0.19
N	8.87 E	2.25 N	25.55	0.09	14.6	25.76	0.07	25.81	0.13	25.71	0.10	25.94	0.20
O	4.06 E	3.11 N	28.61*	0.34	3.2	-	-	-	-	-	-	-	-
P	6.21 E	10.31 S	27.28	0.22	5.1	-	-	-	-	-	-	-	-
Q	5.81 E	10.38 S	27.37*	0.25	4.3	-	-	-	-	-	-	-	-
R	3.29 W	11.57 S	26.99	0.16	7.3	27.04	0.13	27.13*	0.35	27.23*	0.32	27.25*	0.54
S	12.47 E	3.46 S	24.35	0.06	37.3	24.83	0.06	24.21	0.05	24.46	0.05	24.23	0.07
T	12.05 E	1.57 S	26.89	0.21	5.4	26.75	0.20	26.05*	0.25	25.83	0.18	25.84*	0.28
J073149+285449													
A	1.28 E	0.86 S	25.49	0.11	13.3	-	-	-	-	-	-	-	-
B	2.44 E	1.51 S	26.05	0.16	7.4	-	-	-	-	-	-	-	-
C	4.50 E	0.58 S	27.10	0.22	5.2	-	-	-	-	-	-	-	-
D	3.97 W	1.73 S	25.11	0.10	15.0	-	-	-	-	-	-	-	-
E	4.61 W	1.10 N	26.17	0.16	7.4	-	-	-	-	-	-	-	-
F	5.21 W	3.59 N	24.67	0.09	20.2	-	-	-	-	-	-	-	-

Continued on Next Page...

Table 4.6 – Continued

ID	Δ ($''$)		u tot. (mag.)	σ_u (mag.)	S/N$_u$	u ap. (mag.)	$\sigma_{u,c}$ (mag.)	R ap. (mag.)	$\sigma_{R,c}$ (mag.)	V ap. (mag.)	$\sigma_{V,c}$ (mag.)	I ap. (mag.)	$\sigma_{I,c}$ (mag.)
G	8.71 W	5.80 N	27.26*	0.24	4.7	-	-	-	-	-	-	-	-
H	11.19 W	5.98 N	25.13	0.10	14.3	-	-	-	-	-	-	-	-
I	2.44 E	12.14 N	22.86	0.07	64.6	-	-	-	-	-	-	-	-
L	0.26 E	8.59 S	26.39	0.17	6.8	-	-	-	-	-	-	-	-
M	0.58 E	9.90 S	27.02	0.23	5.0	-	-	-	-	-	-	-	-
N	11.81 E	4.63 N	27.72*	0.24	4.7	-	-	-	-	-	-	-	-

4.10.2 Previous imaging studies of high-z DLAs

In this appendix, we review all previous emission studies of $z \gtrsim 2$ DLAs that are known to the authors. Table 4.7 summarizes the properties of the six such DLAs for which a galaxy counterpart has been spectroscopically confirmed. The $z \sim 2.04$ DLA towards PKS0458−02 is included in this list although direct imaging of the associated galaxy is not available (See, however, Warren et al. 2001). For this object, the impact parameter has been computed via long-slit spectroscopy at two orientations. Conversely, the table does not include DLAs whose emission lines have been detected in spectra, but for which no spatial information is available; these are discussed separately below. The columns of Table 4.7 are: (1) The quasar name and a reference to the first identification of the galaxy counterpart. (2) The quasar redshift. (3) The DLA redshift. (4) The impact parameter, in $''$. (5) The impact parameter, in kpc. (6) The H I column density. (7) The published SFR. (8) The SFR diagnostic used for column (7). (9) The unobscured SFR, computed from the UV emission with our calibration, after applying a K-correction and an IGM correction. (10) The Lyα flux in 10^{-17} erg cm^{-2} s^{-1}. (11) The [OIII] flux in 10^{-17} erg cm^{-2} s^{-1}. (12) The broad-band magnitude, in AB mag. (13) The filter used for broad-band photometry in column (12). (14) The half-light radius, deconvolved for the PSF.

Individual references are: (1) Fynbo et al. (1999) (2) Weatherley et al. (2005) (3) Møller et al. (2004) (4) Møller & Warren (1993) (5) Møller & Warren (1998) (6) Wolfe et al. (2005) (7) Møller et al. (2002) (8) Djorgovski et al. (1996) Where not explicitly specified, data are taken from Weatherley et al. (2005) and references therein.

Table 4.7: Properties of DLAs with confirmed galaxy associations. Individual references are given in the main text. References listed with the QSO name refer to the first galaxy-absorber association.

Name	z_{qso}	z_{dla}	b_{as} (")	b_p (kpc)	$N_{\rm HI}$ ($\rm cm^{-2}$)	SFR ($\rm M_\odot\ yr^{-1}$)
(1)	(2)	(3)	(4)	(5)	(6)	(7)
[1]Q0151+048A	1.922	1.934	[1]0.93	7.59	20.36	-
[3]PKS0458-02	2.286	2.039	[3]0.30	2.44	21.65	[6]>1.5
[4]PKS0528-250	2.797	2.811	[5]1.17	8.92	21.35	[4]4.2
[7]Q2206-1958a	2.559	1.920	[7]0.99	8.09	20.65	5.7
[2]Q2206-1958b	2.559	1.920	1.23	10.05	20.65	4.2
[8]2233.9+1318	3.298	3.150	2.51	18.52	20.00	5.9

Name	SFR ($\rm M_\odot\ yr^{-1}$)		F(Lyα) (cgs)	F($[OIII]$) (cgs)	Magnitude	$\rm r_{hl}$ ($''$)
(1)	(8)	(9)	(10)	(11)	(12)-(13)	(14)
[1]Q0151+048A	-	18.45	[1]191	-	[1]22.9 (u)	-
[3]PKS0458-02	Lyα	-	[3]5.4	-	- (-)	-
[4]PKS0528-250	Lyα	3.17	[7]7.4	-	[7]25.43 (V)	[7]0.09
[7]Q2206-1958a	UV	3.23	[6]26	7.6	24.69 (V)	[7]0.5
[2]Q2206-1958b	UV	2.41	-	10.7	25.01 (V)	-
[8]2233.9+1318	UV	2.98	[8]6.4	6.8	25.75 (V)	[7]0.14

Three other DLAs with confirmed redshifts can be found in the literature. A DLA at $z = 3.407$ has been reported to exhibit Lyα emission in the spectrum of PC0953+4749 by Bunker (HST Proposal ID 10437), but no additional imaging has been published. Djorgovski et al. report (private communication in Weatherley et al. 2005) the detection of emission lines from a DLA at $z = 4.1$ towards the quasar DMS2247-0209. The most likely association is a galaxy at $b = 3.3''$, whose inferred SFR is $\sim 0.7\ \rm M_\odot\ yr^{-1}$. Finally, Leibundgut & Robertson (1999) report the detection of Lyα emission which is spatially extended in the absorption trough of a DLA at $z = 3.083$. Due to the large velocity offset between absorption and emission redshifts, the emission feature could result from an object not associated with the DLA.

We next summarize results from other studies that have either identified possible galaxy counterparts of high-z DLAs, or placed upper limits on the host luminosity. Aragon-Salamanca

et al. (1996) reported deep near-IR images of 10 fields containing DLAs, finding two L_k^* DLA candidates at small impact parameters ($\sim 1.2''$). In a long-slit K-band spectroscopic search for Hα emission from eight DLAs at $z > 2$, Bunker et al. (1999) obtained 3σ limits in the range $5.6 - 18h^{-2}$ $M_\odot$ yr^{-1}. Kulkarni et al. (2000; 2001) used HST NICMOS images of two DLAs at $z = 1.892$ and $z = 1.859$ to place 3σ upper limits of $4.0h_{70}^{-2}$ $M_\odot$ yr^{-1} and $1.3h_{70}^{-2}$ $M_\odot$ yr^{-1}, respectively, on the DLA star formation rates. Ellison et al. (2001) found a possible association between the $z = 3.387$ DLA towards the quasar Q0201+113 and a $0.7L^*$ galaxy at an impact parameter of $15h^{-1}$ kpc. In a deep narrow-band imaging survey of six fields with heavy-element quasar absorption lines, Kulkarni et al. (2006) searched for Lyα emission from absorbers at $z = 2.3 - 2.5$, obtaining SFR limits of $\sim 0.9 - 2.7$ $M_\odot$ yr^{-1}, assuming no dust attenuation of the Lyα line. HST photometry for several galaxies detected in 18 DLA fields is presented in Warren et al. (2001). Finally, Christensen et al. (2007) used IFU spectroscopy in the Lyα line to identify candidate hosts for six DLAs at $z > 2$.

Chapter 5

Directly imaging damped Lyman-α galaxies at $z > 2$ – II. The star formation rates of gas rich galaxies

In this chapter, we present preliminary results from Fumagalli et al. (2012), a paper in preparation in which we will discuss the analysis of the imaging survey introduced in Fumagalli et al. (2010b). Our major findings are that, in agreement with previous works (e.g. Wolfe & Chen 2006), damped Lyman$-\alpha$ galaxies exhibit low levels of *in situ* star formation despite being a significant reservoir of neutral hydrogen. Through a statistical analysis of the galaxy distribution in the imaged quasar fields, we further infer that typical damped Lyman$-\alpha$ systems arise either from $\lesssim 0.1L_*$ galaxies in isolation or from satellites in the halos of $\gtrsim 0.1L_*$ galaxies. The incidence of damped Lyman$-\alpha$ systems in extended gas disks around Lyman break galaxies with star formations rates $> 1\mathrm{M}_\odot$ yr^{-1} appears to be low.

5.1 Abstract

We present results from an imaging survey of 32 quasar fields with known damped Lyman-α (DLA) systems at $z \sim 1.9 - 3.8$. To avoid the glare from the bright background quasar that hampered previous studies, we select sightlines which have optically thick absorbers at higher redshifts. These systems fully block the quasar light in the rest-frame FUV of the targeted DLA. In 2/32 fields, we detect emission at the projected position of the quasar that we interpret as *in situ* star formation from the DLA galaxies with star formation rates (SFRs) of 0.9 $M_\odot$ $yr^{-1}$and 2.7 $M_\odot$ yr^{-1}. For the remaining 30 sightlines, instead, we place upper limits at ~ 1 $M_\odot$ yr^{-1} for both extended (10 kpc) and compact (3 kpc) sources. The corresponding limits on the star formation surface densities are $< 0.01 - 0.03$ $M_\odot$ yr^{-1} kpc^{-2}. The lack of appreciable *in situ* SFR in DLAs is expected on theoretical grounds because of the low hydrogen column density and low metal content. Further, we study the environment around these quasar sightlines, finding that $> 90\%$ ($> 97\%$) of the galaxies with SFRs between $1 - 5$ $M_\odot$ yr^{-1}($5 - 25$ $M_\odot$ yr^{-1}) lie beyond 10 kpc from the projected quasar location and that the occurrence of very bright galaxies (SFR$>$ 25 $M_\odot$ yr^{-1}; $\gg L_*$) galaxies in the inner 100 kpc is rare. We conclude that either DLAs arise from faint $\lesssim 0.1L_*$ galaxies that are centrals to their halos or from clumps and satellites in the halo of galaxies up to a few L_*, but not in the outskirts ($<$ 10kpc) of galaxies with SFR$>$ 1 $M_\odot$ yr^{-1}.

5.2 Introduction

The continuous improvement of instrumentation available at ground based observatories and on board of the *Hubble Space Telescope* (HST) has enabled numerous deep galaxies surveys on wide areas that characterize the typical star formation rates (SFR), dust content, stellar masses and structural properties of galaxies across cosmic time, up to the end of the

reionization epoch. However, the distance dependent sensitivity limits of these surveys imply that optically-selected samples probe only the tip of the iceberg of the galaxy population in the distant universe.

An appealing alternative to probe the physical properties of galaxies across a wider range of luminosity and masses is the study of damped Lyman-α (DLA) systems (Wolfe et al. 2005; and references therein) that are detected in absorption along the line of sight to bright background sources such as quasars (QSOs) or the afterglows of γ-ray bursts. In fact, being selected purely based on their neutral hydrogen column density ($N_{\rm HI} > 2 \times 10^{20}$ cm^{-2}), DLAs arise in dense and neutral gas and, as suggested by models and hydrodynamic simulations, trace the hydrogen content of galaxies over a wide range of halo masses (e.g. Nagamine et al. 2007, Pontzen et al. 2008, Tescari et al. 2009, Cen 2012, Hong et al. 2010, Fumagalli et al. 2011a).

The study of the hydrogen absorption lines in large spectroscopic samples of DLAs (e.g. Prochaska & Wolfe 2009, Noterdaeme et al. 2009) and of the associated metal lines in high-resolution spectra (e.g Prochaska et al. 2007b, Rafelski et al. 2012), as well as the detailed analysis of the abundance patterns (e.g. Dessauges-Zavadsky, Calura, Prochaska et al. 2007), have allowed a precise characterization of the physical state of the gas in these absorbers at different redshifts. DLAs cover roughly one third of the sky over the interval $z = 2 - 3$ and evolve in time such that the shape of the column density distribution function is preserved. They are typically sub-solar in composition, but more enriched ($\sim 10\%$ of the solar metallicity) compared to the intergalactic medium (IGM). Finally, DLAs exhibit abundance patterns that can be described with bursty star formation histories common to irregular and dwarf galaxies, or, as suggested by more recent results, that follows those observed in the Galactic halo (Rafelski et al. 2012).

An outstanding challenge that limits a definitive understanding of the nature of DLAs and thus the link to the population of optically selected galaxies is the ability to connect the

absorption systems to their emitting counterparts. In fact, the bright background light that enables spectroscopy becomes an inconvenient source of contamination during imaging observations, precluding the identification of faint galaxies at close impact parameters. For this reason, despite over twenty years of investigation (see a review of previous searches in Fumagalli et al. 2010b), the emitting counterparts of only 10 DLAs have been detected to date at $z \gtrsim 2$ (for the most recent compilation, see Krogager et al. 2012). This number has rapidly increased only in recent years thanks to new spectroscopic and integral field unit surveys (Fynbo et al. 2010; 2011, Péroux et al. 2012). While these efforts are providing the first glimpes at the emission properties of DLA galaxies, the samples studied to date are highly heterogeneous and perhaps offer only a biased representation of the DLA population.

Pressing questions on what the typical sizes of DLAs are or at what rate they form stars remain open. Are DLA galaxies massive rotating disks as suggested by the absorption line profiles (Prochaska & Wolfe 1997) or the small building blocks predicted by the theories of the hierarchical assembly of structures (Haehnelt et al. 1998, Rauch et al. 2008)? And what is the source of heating that balances the cooling rates observed via the [C II*] absorption lines (Wolfe et al. 2003)? Or, if the *in situ* star formation rates (SFRs) are too low to fully account for the chemical enrichment and number density evolution of DLAs (Wolfe & Chen 2006, Prochaska & Wolfe 2009), are these systems associated to the outskirts of more compact Lyman break galaxies (LBGs) (Rafelski et al. 2011)? To address these and similar questions we have undertaken an imaging survey of 32 QSO fields with intervening DLAs at $z \sim 1.9 - 3.8$, taking advantage of an observational technique that prevents the contamination from the bright background source (O'Meara et al. 2006).

With this survey, we aim to constrain three possible scenarios for the origin of DLAs. The first scenario we wish to examine is one in which the majority of DLAs arise from a gaseous disk in the outskirt of star-forming galaxies. In this picture, DLAs are the fuel for star formation

at low surface density and, perhaps, the reservoir for future star formation in the inner disk. If this is the case, we do not expect to detect emission at the DLA position at the depth of our imaging, yet we should expect to see brighter star-forming galaxies within $\sim 10-15$ kpc, i.e. a reasonable size for an H I disk at high redshift (see Paper I, figure 13). A second scenario we consider is one in which DLAs form from gas that is physically associated to a star-forming galaxy, but not within a gaseous disk that extends from the central galaxy. Rather, these systems form in satellites or dense gas clumps in the halo of the central galaxy. Again, if this is the case, we should not expect to detect emission at the DLA location, but from a galaxy that is most likely located beyond $\sim 10-15$ kpc from the projected location of the QSO, up to the virial radius of the central galaxy (~ 100 kpc for massive halos). Finally, in a third scenario, DLAs arise in faint galaxies with low SFRs and that are centrals to their halos. In this case, we should expect to detect only a limited number of DLA counterparts in emission. Further, we should not frequently see star-forming galaxies in the surroundings.

The survey design and observations of two fields have been presented in the first paper of this series (Fumagalli et al. 2010b; hereafter Paper I). In this paper, after a brief review of the adopted technique and of the sample selection (Section 5.3), we present ground based and HST imaging observations for the 32 QSO fields, together with new echellette spectra of 18 QSOs (Section 5.4). The first unbiased census of the *in situ* SFRs of DLAs and a discussion on the nature of DLAs inferred from the environment in proximity to the projected quasar location are presented in Section 5.5. A summary and conclusions follow in Section 5.6. This paper is complemented by Appendix 5.7 in which we present tables of the ionic column densities in 17 DLAs with unpublished metallicity. Throughout this paper, we adopt the WMAP7 cosmological parameters from Komatsu et al. (2011).

5.3 Technique and sample selection

To overcome the glare of the background QSOs that would prevent the detection of faint galaxies at small projected separations, we select QSO fields that host an additional optically thick absorbers along the line of sight. This technique, firstly presented by O'Meara et al. (2006), is detailed in Paper I and here only briefly summarized.

The basic idea is to select quasar fields in which there are optically-thick absorbers (which we dub 'blocking filters') at distances greater than those to the targeted DLAs. These redshift separations are carefully chosen such that the corresponding Lyman limit[1] of the blocking filters lie redward to the $U-$ or $B-$band filters available at optical telescopes. At the same time, the different distances are chosen to allow the detection of the Lyα line of the targeted DLAs. These natural blocking filters absorb the QSO light allowing direct imaging of the rest-frame FUV light of galaxies potentially associated to the lower-redshift DLAs at all impact parameters.

Among the known DLAs at $z < 4$, there are ~ 100 sightlines that can be imaged with ground based telescopes or HST with this technique. Our final sample is composed by 32 QSO sightlines that are further divided in two sub-samples. The HST sub-sample encompasses 20 sightlines that were selected to match the filters of the WFC3/UVIS camera, while the Keck sample is composed by 12 sightlines that were selected to match the U-band and $B-$band filters at Keck/LRIS and LBT/LBC and that were visible in the scheduled observing nights. No other selection criteria (e.g. column density of metallicity) were adopted. Table 5.1 and 5.2 summarize the basic properties of the observed QSO fields and of the targeted DLAs.

In Figure 5.1, we show the distribution of redshifts, metallicities and column densities of the observed DLAs (see Section 5.4 for details on how these quantities were derived). As a result

[1]The Limit limit is the sharp cutoff in the transmitted quasar light due to a complete absorption of the ionizing radiation in the rest frame of an intervening optically thick system

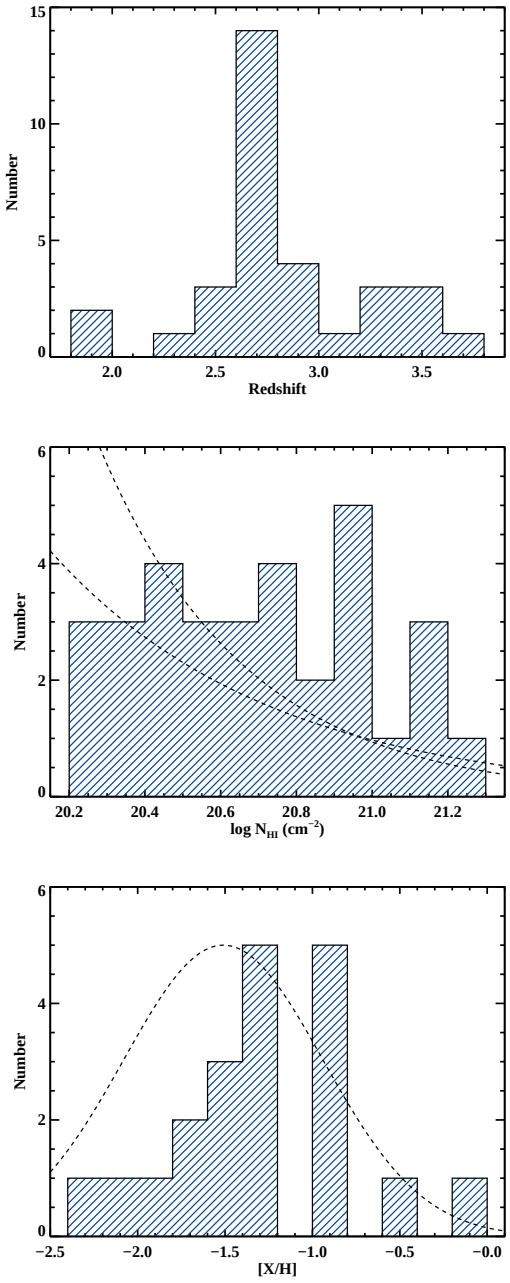

Figure 5.1: Histograms of the redshift (top), H I column density (middle), and metallicity (bottom) for the selected DLAs. In the middle panel, we superimpose the column density distribution functions (dashed lines) from Prochaska & Wolfe (2009) in the redshift interval [2.4,2.7] and [2.7,3.0]. In the bottom panel, instead, we overplot the metallicity distribution for DLAs at $z > 1.5$ from Rafelski et al. (2012). Being assembled purely based on the presence of multiple absorbers in the line of sight, our sample is representative of the DLA population at $z \sim 2.7$. However, due to the small number of systems, the resultant column densities and metallicities are slightly skewed to large values.

of the adopted selection criteria, the redshift distribution (top panel) peaks at $z \sim 2.7$. Since no other constraints were applied, our survey should be representative of the DLA population at $z \sim 2 - 3$. In fact, the metallicity distribution (bottom panel) ranges from $[X/H] \sim -2.5$ to nearly solar values, with a peak at $[X/H] \sim -1$. This is slightly skewed to higher metallicity than the distribution observed in larger DLA samples (e.g. Prochaska et al. 2007b, Rafelski et al. 2012). Similarly, the column density distribution (central panel) appears slightly offset towards high values, probably due to the limited number of sightlines.

Table 5.1: Log book of the ground-based and HST imaging observations and summary of the quasar field properties.

Field	R.A. (J2000)	Dec. (J2000)	UT Date	Filter	Instr.	Exp. Time (s)	FWHM ($''$)	3σ depth (mag in $1''$ ap.)	E(B-V) (mag)
1:K1	21:14:43.95	-00:55:32.7	2008 Oct, 2nd	u	LRIS	5400	0.6	28.34	0.062
			”	V	LRIS	1320	0.5	27.61	
			”	R	LRIS	1320	0.6	27.06	
			”	I	LRIS	1515	0.7	26.73	
2:K2	07:31:49.50	+28:54:48.6	2009 Jan, 28th	u	LRIS	5400	0.8	28.02	0.055
			”	V	LRIS	2140	0.8	26.95	
			”	R	LRIS	2420	0.7	26.43	
3:K3	09:56:04.43	+34:44:15.5	2009 Jan, 28th	u	LRIS	5400	0.9	28.59	0.013
			”	V	LRIS	2280	0.8	27.64	
			”	R	LRIS	2280	0.8	27.13	
4:K4	23:43:49.41	-10:47:42.0	2009 Jul, 23rd; 2009 Sep 17th	u	LRIS	6300	1.2	28.27	0.033
			”	V	LRIS	1400	0.9	27.12	
			”	R	LRIS	1440	1.1	26.79	
			”	I	LRIS	1540	0.9	26.62	
5:K5	03:43:00.88	-06:22:29.9	2009 Sep, 20th	Us	LBC	3900	1.1	27.79	0.060
			”	V	LBC	900	0.9	27.24	
			”	R	LBC	1200	0.8	27.03	
6:K6	23:51:52.80	+16:00:48.9	2009 Sep, 21st	B	LBC	1650	1.1	27.96	0.037
			”	V	LBC	450	0.9	26.92	
			”	R	LBC	600	0.9	26.75	
			”	I	LBC	750	0.9	26.80	
7:K7	00:42:19.74	-10:20:09.4	2009 Dec, 17st	u	LRIS	5400	0.7	28.55	0.033
			”	V	LRIS	1400	0.7	27.28	
			”	R	LRIS	1400	0.7	26.79	
			”	I	LRIS	1440	0.7	26.64	
8:K9	09:49:27.88	+11:15:18.2	2009 Dec, 17st	u	LRIS	5400	0.7	28.74	0.024
			”	V	LRIS	1200	0.8	27.35	
			”	R	LRIS	1200	0.7	27.06	

Continued on Next Page...

Table 5.1 – Continued

Field	R.A. (J2000)	Dec. (J2000)	UT Date	Filter	Instr.	Exp. Time (s)	FWHM ($''$)	3σ depth (mag in $1''$ ap.)	E(B-V) (mag)
			”	I	LRIS	1440	0.8	26.95	
9:K10	10:18:06.28	+31:06:27.2	2009 Dec, 17st	u	LRIS	3600	0.8	28.56	0.031
			”	V	LRIS	800	0.8	26.92	
			”	R	LRIS	800	0.8	26.77	
			”	I	LRIS	960	0.8	26.75	
10:K11	08:51:43.72	+23:32:08.9	2009 Dec, 18st	B	LRIS	3600	1.1	28.06	0.033
			”	V	LRIS	800	1.1	26.80	
			”	R	LRIS	800	1.1	26.42	
			”	I	LRIS	960	0.9	26.26	
11:K12	09:56:05.09	+14:48:54.7	2009 Dec, 18st	U	LRIS	5400	1.1	28.64	0.027
			”	V	LRIS	1200	1.1	27.30	
			”	R	LRIS	1200	1.0	26.95	
			”	I	LRIS	1440	1.0	26.78	
12:K13	11:51:30.48	+35:36:25.0	2009 Dec, 18st	U	LRIS	3600	1.1	28.67	0.020
			”	V	LRIS	1200	1.0	27.38	
			”	R	LRIS	1200	1.0	27.17	
			”	I	LRIS	1440	0.8	27.06	
13:H1	21:23:57.56	-00:53:50.1	2009 Dec, 4th	$F390M$	UVIS	5130	0.08	25.43	0.039
14:H2	04:07:18.20	-44:10:14.0	2010 Feb, 13th	$F275W$	UVIS	5361	0.09	26.17	0.013
15:H3	02:55:18.58	+00:48:47.6	2010 Feb, 13th	$F390W$	UVIS	5130	0.08	26.79	0.087
			2009 Dec, 17st	u	LRIS	4500	0.7	28.41	
			”	V	LRIS	1000	0.6	27.01	
			”	R	LRIS	1000	0.7	26.76	
			”	I	LRIS	1200	0.6	26.50	
16:H4	08:16:18.99	+48:23:28.4	2010 Feb, 24th	$F390M$	UVIS	5456	0.08	25.53	0.054
			2012 Jan, 25th	V	ESI	360	0.9	27.06	
			”	R	ESI	360	0.9	26.20	
17:H5	09:30:51.93	+60:23:01.1	2010 Feb, 28th	$F390M$	UVIS	5721	0.08	25.32	0.031
			2012 Mar, 17th	V	ESI	360	0.8	26.83	
			”	R	ESI	360	0.7	26.23	

Continued on Next Page...

Table 5.1 – Continued

Field	R.A. (J2000)	Dec. (J2000)	UT Date	Filter	Instr.	Exp. Time (s)	FWHM ($''$)	3σ depth (mag in $1''$ ap.)	E(B-V) (mag)
18:H6	09:08:10.36	+02:38:18.7	2010 Mar, 20$^{\rm th}$	$F390M$	UVIS	5641	0.08	25.62	0.027
			2012 Jan, 25$^{\rm th}$	V	ESI	360	1.0	26.90	
			"	R	ESI	360	0.8	26.20	
19:H7	12:20:21.39	+09:21:35.7	2010 Apr, 13$^{\rm th}$	$F438W$	UVIS	5657	0.08	26.36	0.025
			2012 Jan, 25$^{\rm th}$	V	ESI	360	0.8	26.95	
			"	R	ESI	360	0.8	26.29	
20:H8	14:42:33.01	+49:52:42.6	2010 Apr, 16$^{\rm th}$	$F343N$	UVIS	5955	0.08	26.09	0.029
			2012 Mar, 17$^{\rm th}$	V	ESI	360	0.8	26.75	
			"	R	ESI	360	0.7	26.30	
21:H9	08:44:24.24	+12:45:46.7	2010 Apr, 24$^{\rm th}$	$F275W$	UVIS	5663	0.09	26.32	0.045
			2012 Jan, 25$^{\rm th}$	V	ESI	360	0.8	26.75	
			"	R	ESI	360	0.7	26.19	
22:H10	07:51:55.10	+45:16:19.6	2010 May, 4$^{\rm th}$	$F343N$	UVIS	5955	0.08	25.76	0.049
			2012 Jan, 25$^{\rm th}$	V	ESI	360	0.9	26.96	
			"	R	ESI	360	0.8	26.30	
23:H11	08:18:13.14	+07:20:54.9	2010 May, 10$^{\rm th}$	$F390W$	UVIS	5657	0.08	26.73	0.019
			2012 Jan, 25$^{\rm th}$	V	ESI	360	0.8	26.90	
			"	R	ESI	360	0.8	26.15	
24:H12	08:18:13.05	+26:31:36.9	2010 May, 13$^{\rm th}$	$F438W$	UVIS	5687	0.08	26.14	0.032
			2012 Jan, 25$^{\rm th}$	V	ESI	360	0.9	26.84	
			"	R	ESI	360	0.8	26.13	
25:H13	08:11:14.32	+39:36:33.2	2010 May, 14$^{\rm th}$	$F343N$	UVIS	5775	0.08	25.82	0.050
			2012 Jan, 25$^{\rm th}$	V	ESI	360	0.9	26.96	
			"	R	ESI	360	0.8	26.21	
26:H14	15:08:51.94	+51:56:27.7	2010 Jun, 1$^{\rm st}$	$F390M$	UVIS	6063	0.08	25.52	0.021
			2012 Mar, 17$^{\rm th}$	V	ESI	360	0.9	26.77	
			"	R	ESI	360	0.8	26.23	
27:H15	10:54:30.07	+49:19:47.1	2010 Jun, 4$^{\rm th}$	$F390M$	UVIS	5955	0.08	25.53	0.020
			2012 Mar, 17$^{\rm th}$	V	ESI	360	1.0	27.05	
			"	R	ESI	360	0.9	26.40	

Continued on Next Page...

Table 5.1 – Continued

Field	R.A. (J2000)	Dec. (J2000)	UT Date	Filter	Instr.	Exp. Time (s)	FWHM ($''$)	3σ depth (mag in $1''$ ap.)	E(B-V) (mag)
28:H16	09:56:25.16	+47:34:42.5	2010 Jun, 6$^{\rm th}$	$F438W$	UVIS	5955	0.08	26.19	0.010
			2012 Mar, 17$^{\rm th}$	V	ESI	360	0.7	26.95	
			”	R	ESI	360	0.8	26.33	
29:H17	14:41:47.52	+54:15:38.1	2010 Jun, 9$^{\rm th}$	$F343N$	UVIS	6063	0.08	26.14	0.014
			2012 Jan, 25$^{\rm th}$	V	ESI	360	1.1	26.75	
			”	R	ESI	360	0.9	26.26	
30:H18	11:55:38.60	+05:30:50.5	2010 Jul, 12$^{\rm th}$	$F343N$	UVIS	5657	0.08	25.81	0.018
			2012 Jan, 25$^{\rm th}$	V	ESI	360	0.8	26.94	
			”	R	ESI	360	0.7	26.23	
31:H19	15:24:13.35	+43:05:37.4	2010 Jul, 14$^{\rm th}$	$F390W$	UVIS	5855	0.08	27.00	0.025
			2012 Mar, 17$^{\rm th}$	V	ESI	360	0.8	26.73	
			”	R	ESI	360	0.8	26.24	
32:H20	13:20:05.97	+13:10:15.3	2010 Nov, 23$^{\rm rd}$	$F336W$	UVIS	5663	0.08	26.33	0.026
			2012 Jan, 25$^{\rm th}$	V	ESI	360	0.8	27.04	
			”	R	ESI	360	0.9	26.42	

5.4 Observations

5.4.1 Imaging observations

Among the 12 QSO sightlines in the Keck sub-samples, 10 fields were imaged with Keck/LRIS during seven observing nights allocated between 2008 and 2009. The remaining two fields were observed with LBT/LBC in September 2009. Since both instruments allow simultaneous observations in a blue and a red sensitive camera, we adopted the following observing strategy. For each field, we acquired a set of exposures in the blue sensitive channel either using the $U-$band or $B-$band filters, according to the redshift of the DLAs and of the blocking filters. At the same time, we used the red sensitive camera to image the same fields with multiple filters (typically R, V, and I). The HST targets were instead imaged with a single UV filter carefully matched to the absorber redshifts with the UVIS channel of the recently installed WFC3. Observations were completed in 40 orbits that were allocated in cycle 17 (PI O'Meara, #11595). In support of the HST observations, we also acquired Keck/ESI imaging in the $R-$ and $V-$ band using Keck/ESI. A log book that summarizes the date of the observations, the exposure times, and the instrument and filter used for each field is in Table 5.1.

Since at the completion of our survey, the HST imaging was already available through the Hubble Legacy Archive, we retrieved the already combined images with no further processing. Meanwhile, we reduced the ground based imaging following standard techniques. First, after subtracting the bias, we flat fielded each exposure using dome flats for the red sensitive camera and twilight flats for the blue sensitive one. We then converted counts in units of electron/second and we fit an astrometric solution to each frame using the IRAF task `ccdmap` and reference stars from the Sloan Digital Sky Survey (Abazajian et al. 2009; SDSS-DR7). Then, using the `scamp` software (Bertin 2006), we corrected the astrometric solution and we projected individual exposures to a common grid with pixel size of $0.135''$/pixel for LRIS, $0.224''$/pixel for LBC and

$0.156''$/pixel for ESI. Finally, we combined multiple exposures of the same field using the `swarp` software (Bertin et al. 2002) by means of a weighted mean. The weight images were constructed to avoid bad columns, to correct for pixels with limited sensitivity, and to reject cosmic rays. A gallery of the final images is in Figure 5.2.

For the ground based images, we estimate the image quality by measuring the full-width at half-maximum (FWHM) of point sources. Using `SExtractor` (Bertin & Arnouts 1996), we iteratively measure the FWHM of sources that are classified as stars based on the star parameter (> 0.98) and ellipticity (< 0.15) until the mode of the distribution is converged. These values, listed in Table 5.1, reflect the effective seeing of the images, i.e. they take into account both for the convolution of the different seeings in individual exposures and possible smearing that occurs in the alignment process. For the HST images, instead, we adopt a constant value of $0.09''$ for the F275W filter and $0.08''$ for all the others.

To generate the source catalogue, we run `SExtractor` with a minimum area of 22 pixels and a deblending parameter of 0.02 for the ESI data, a minimum area of 25 pixels and deblending parameter of 0.02 for the LRIS images, and a minimum area of 15 pixels and deblending parameter of 0.01 for the LBC data. In all the cases, we assume a low threshold of 1.2σ. This combination of parameters still allows the inclusion of compact objects in the final catalogue, but it is less sensitive to sharp objects such as residual cosmic rays. For the HST imaging, instead, we use the parameters adopted in the WFC3 Early Release Science observations (Windhorst et al. 2011). Throughout this work, we quote AB magnitudes that are derived in Kron apertures.

Since the ground based imaging has been acquired with multiple instruments and in different weather conditions that ranged from clear sky to slightly overcast, we homogenize these data to a common photometric system. We calibrate our photometry against SDSS by matching the $U-$band filters to the SDSS $u_{\rm sdss}$, the $B-$ and $V-$ band filters to the $g_{\rm sdss}$, and the $R-$

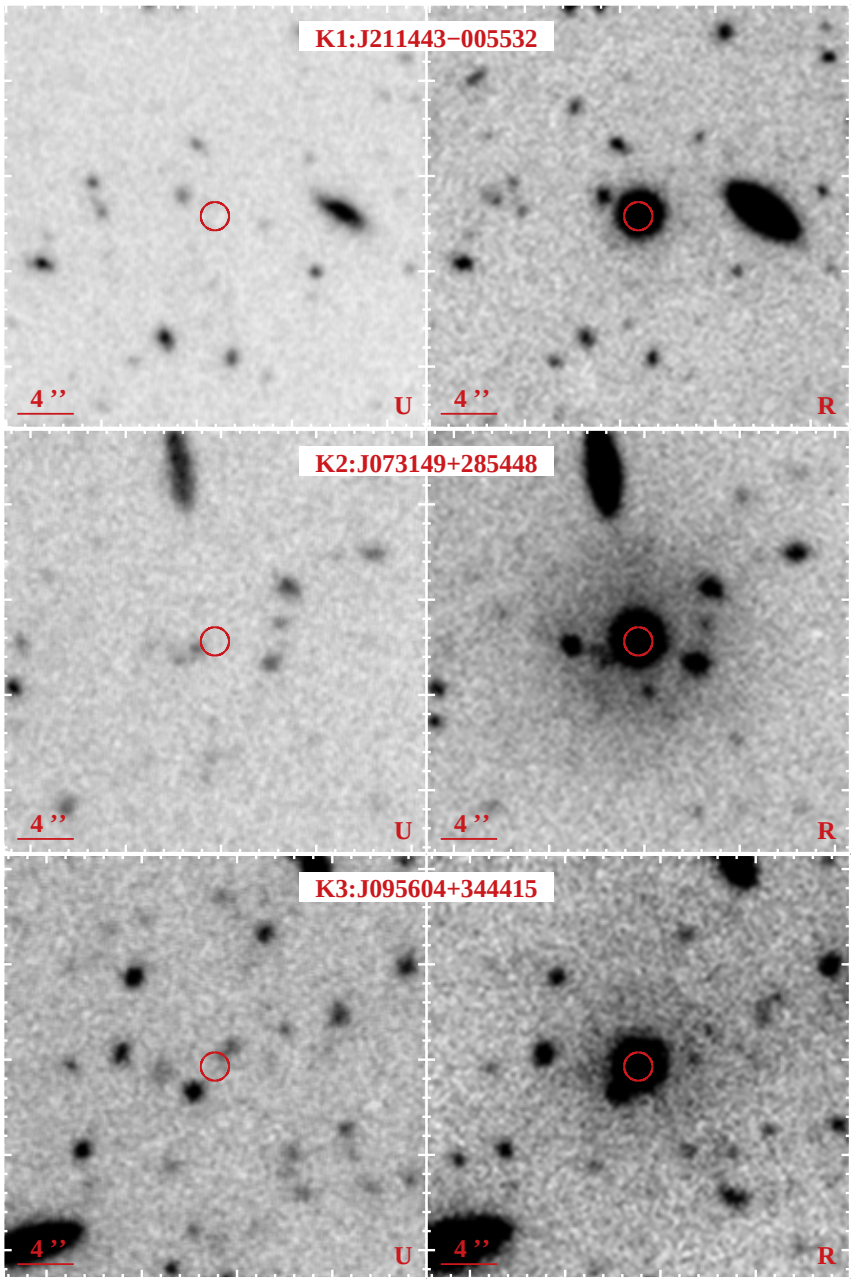

Figure 5.2: Gallery of the imaged fields. For each quasar sightline we show the $U-$ or $B-$band image (left) and the corresponding $R-$band image. A circle marks the quasar position. Due to the optically thick intervening system that acts as a natural blocking filter, the quasar rest-frame FUV light is visible only in the observed $R-$band. Thus, faint galaxies can be detected at any impact parameter without contamination from the background source. The halos visible around the quasars of DLA K2 and K3 are due to condensation on the LRIS camera (see Paper I).

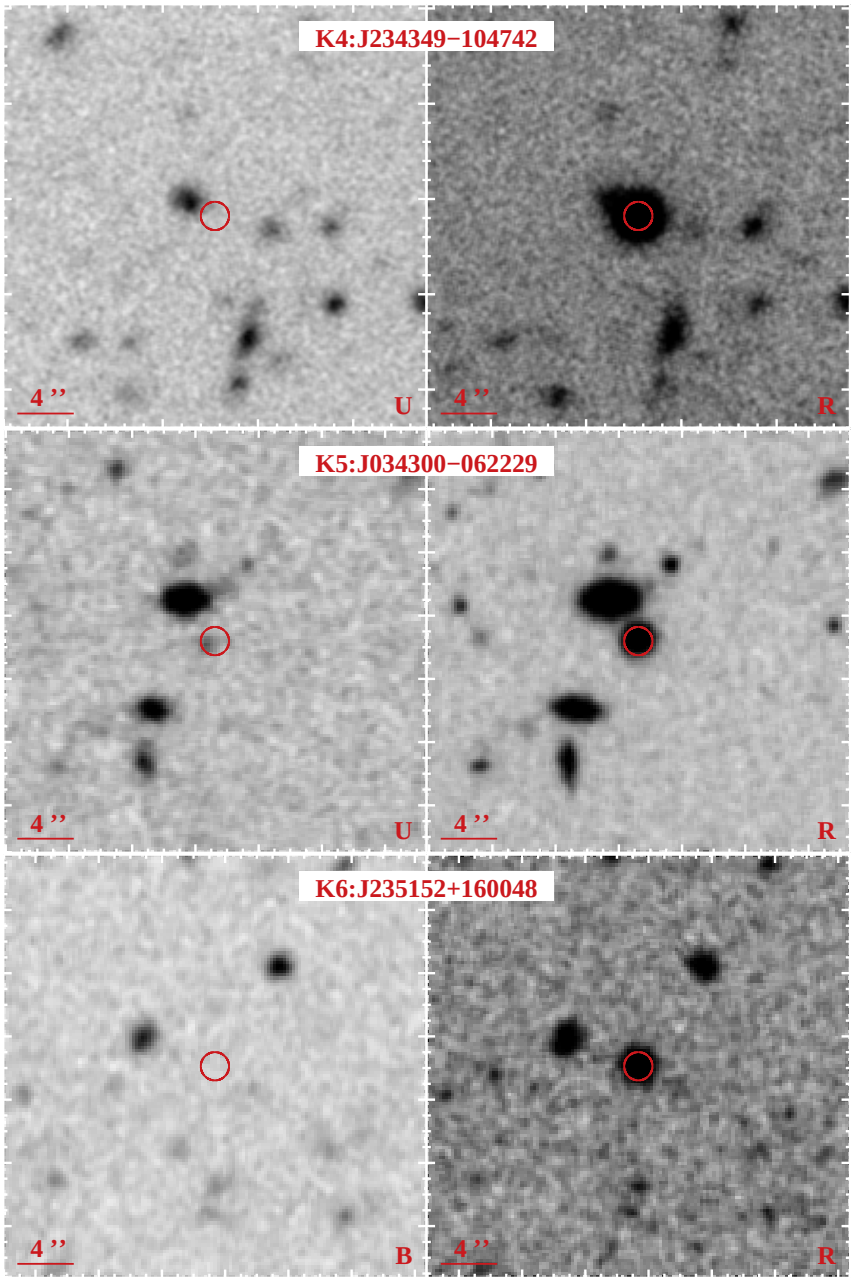

Figure 5.3: Gallery of the imaged quasar fields (continued).

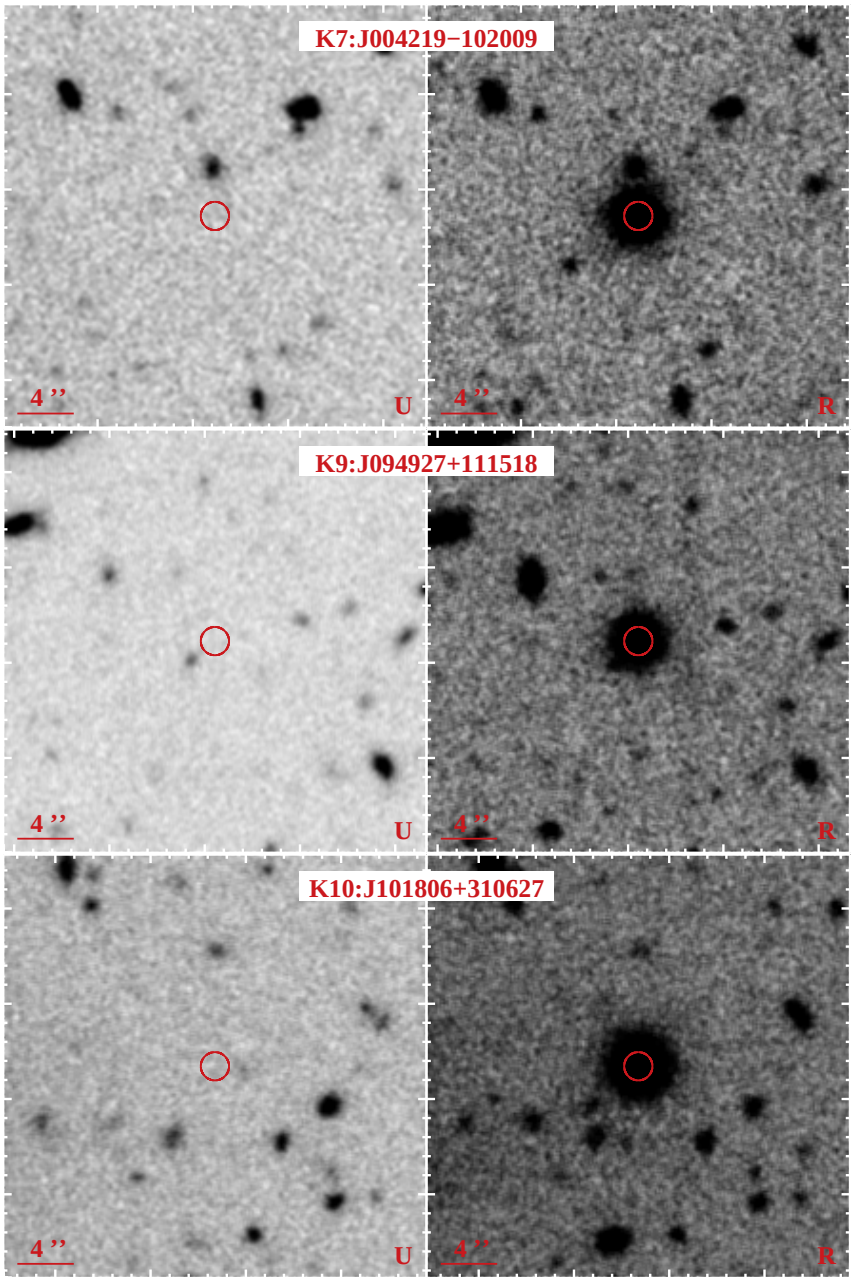

Figure 5.4: Gallery of the imaged quasar fields (continued).

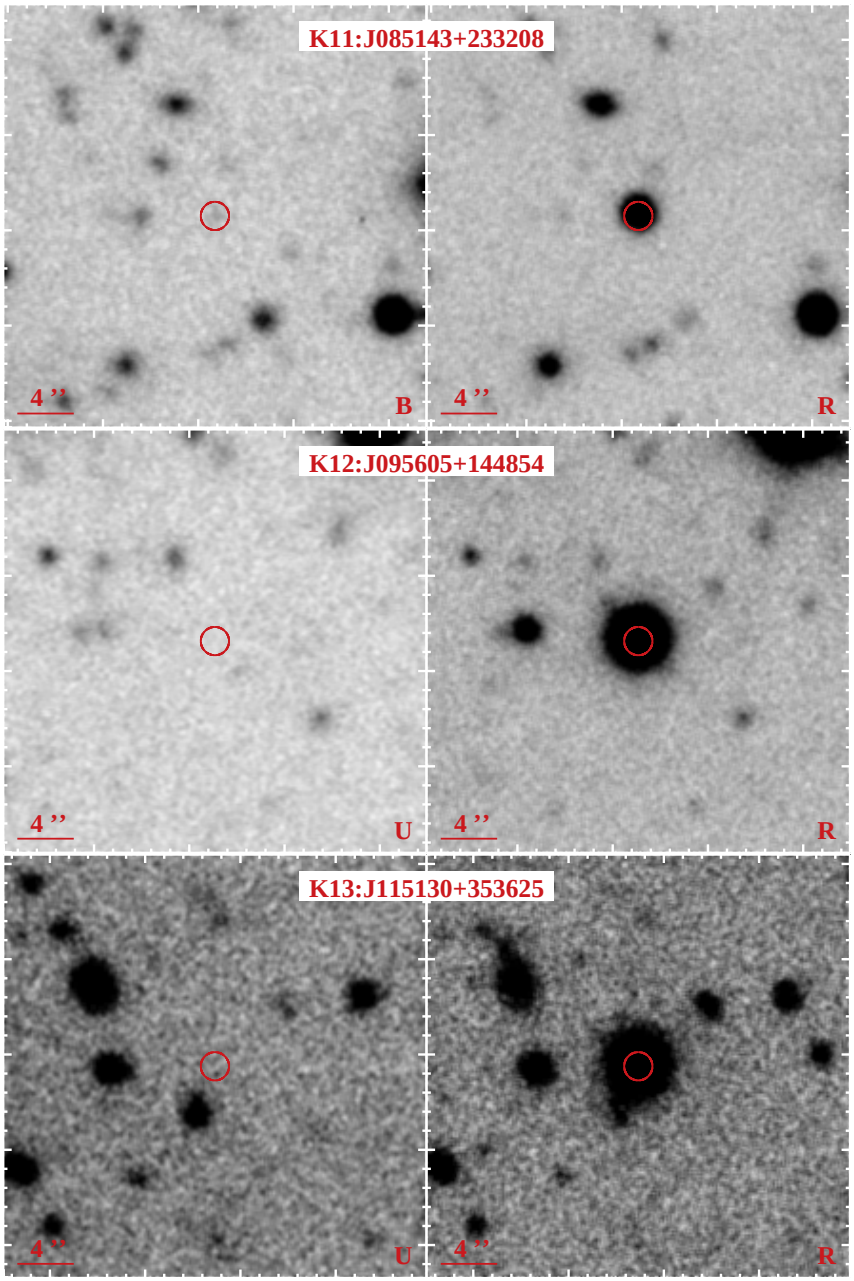

Figure 5.5: Gallery of the imaged quasar fields (continued).

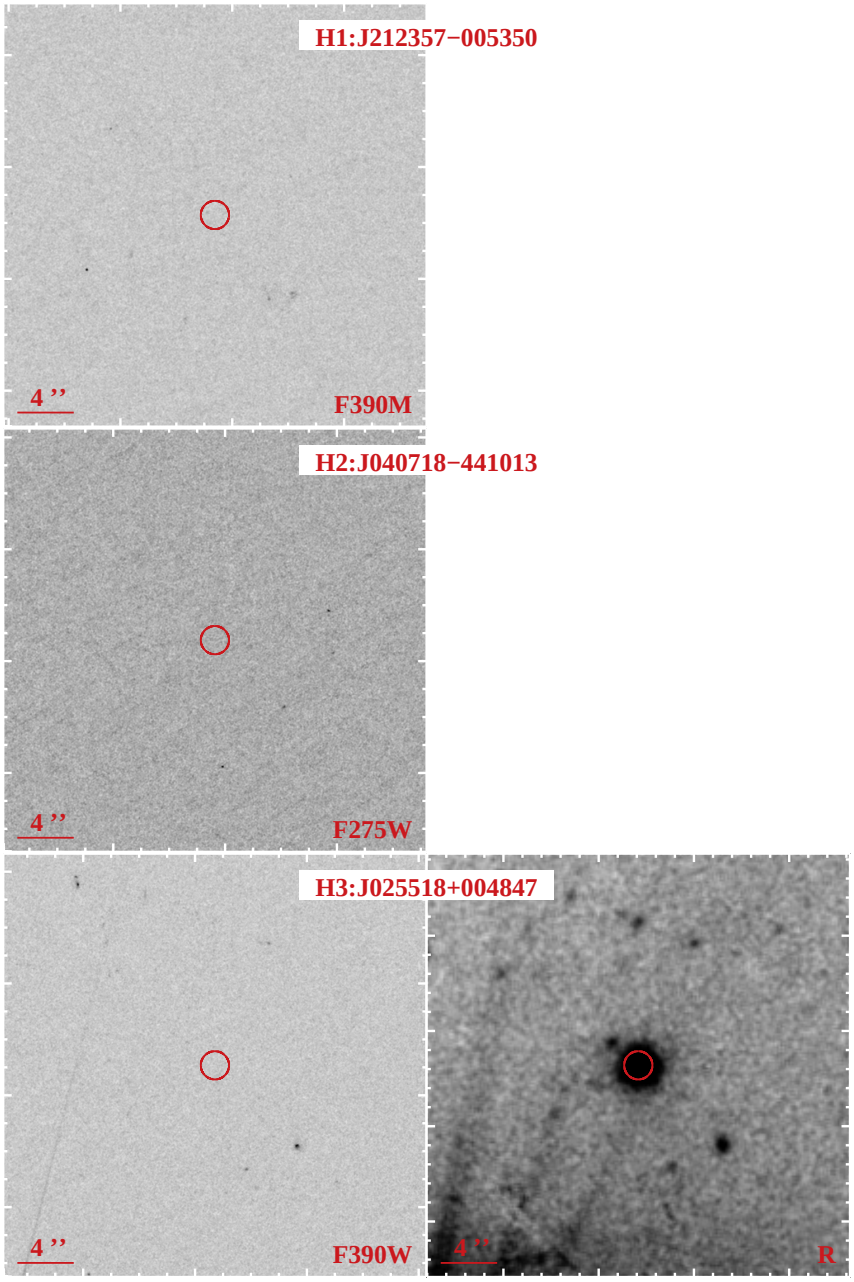

Figure 5.6: Gallery of the imaged quasar fields (continued). The $R-$band images for the H1 and H2 fields are not available.

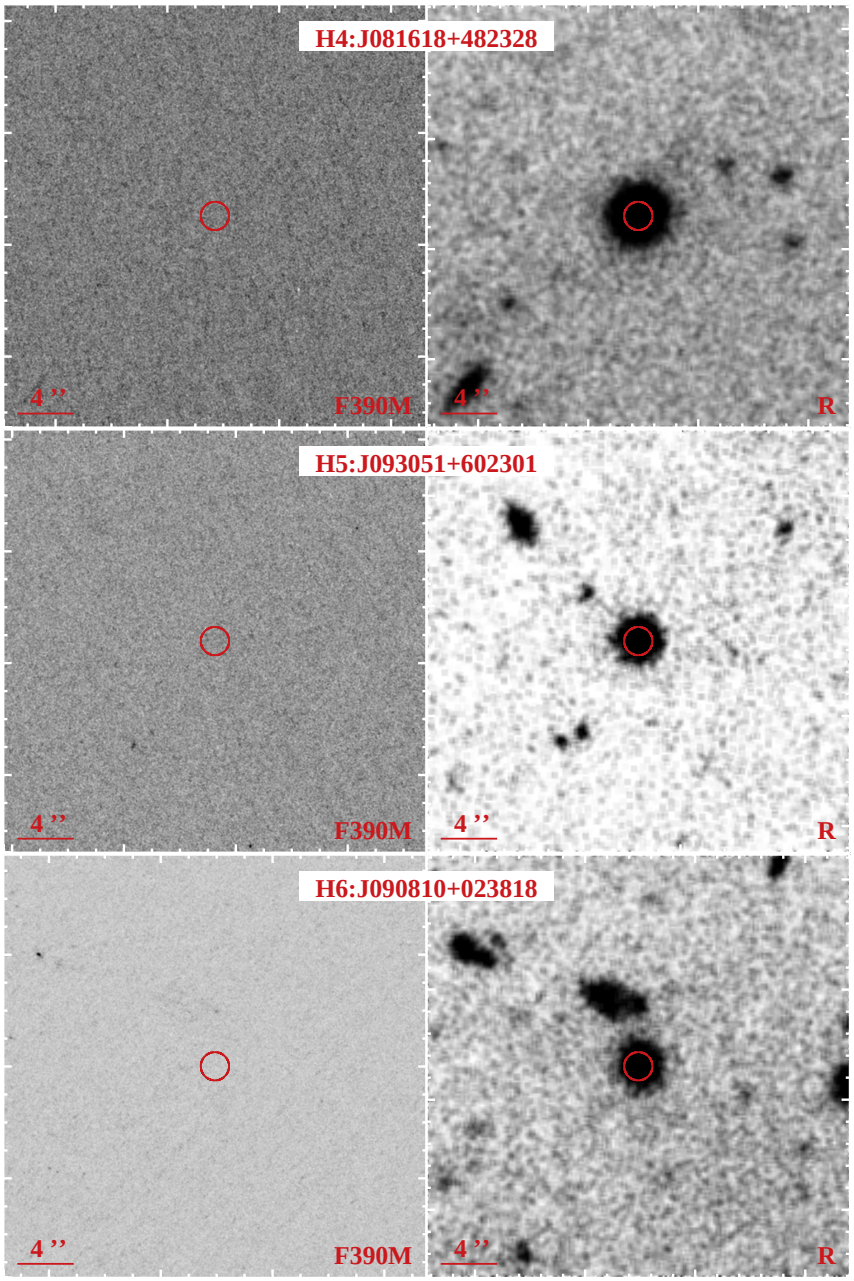

Figure 5.7: Gallery of the imaged quasar fields (continued).

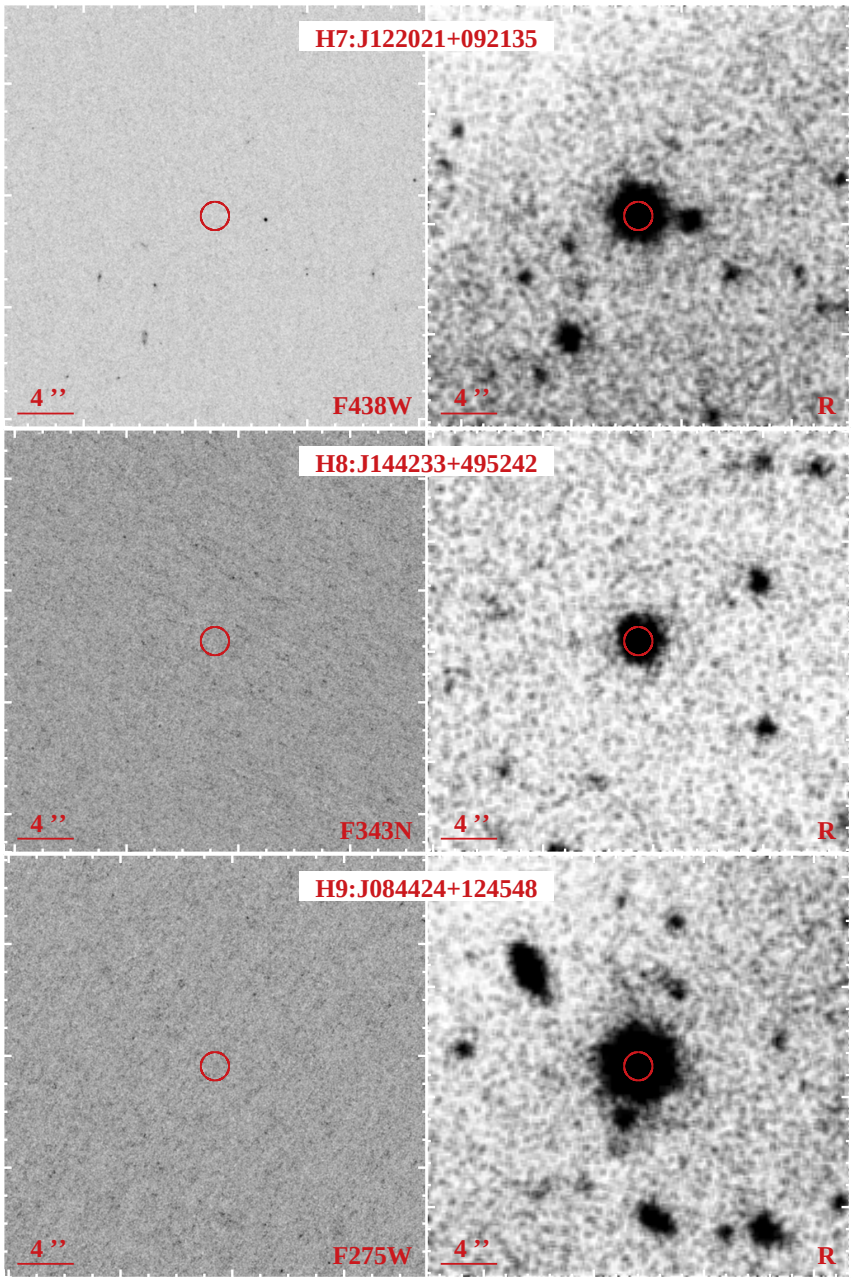

Figure 5.8: Gallery of the imaged quasar fields (continued).

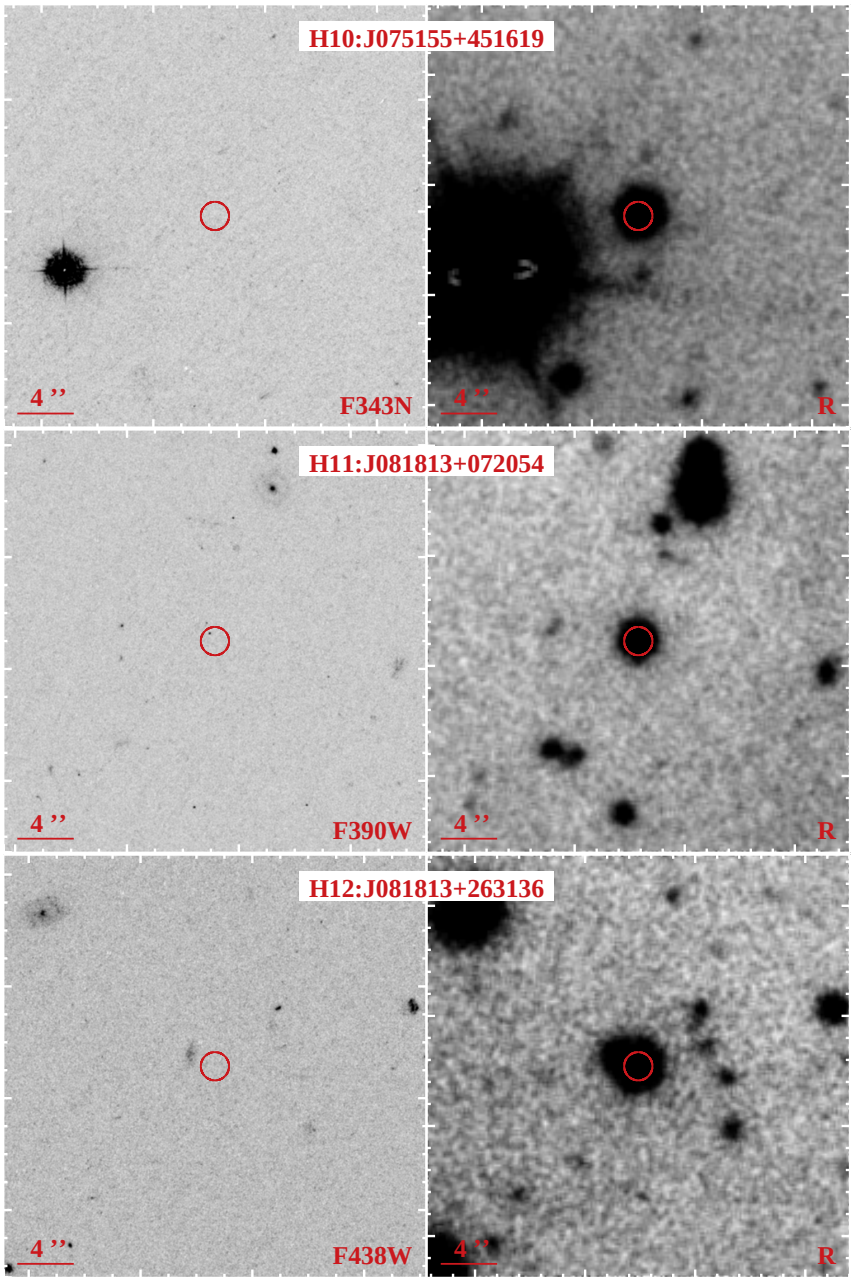

Figure 5.9: Gallery of the imaged quasar fields (continued).

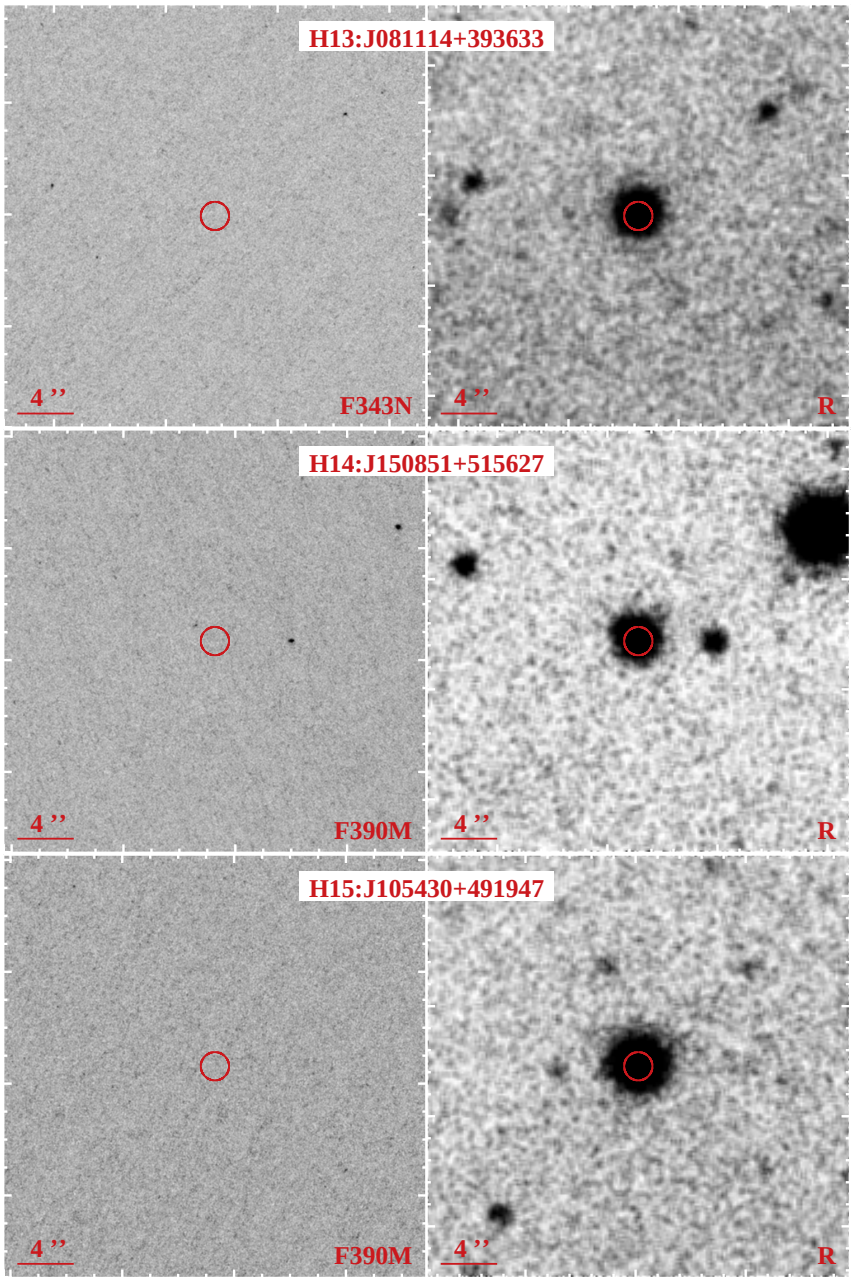

Figure 5.10: Gallery of the imaged quasar fields (continued).

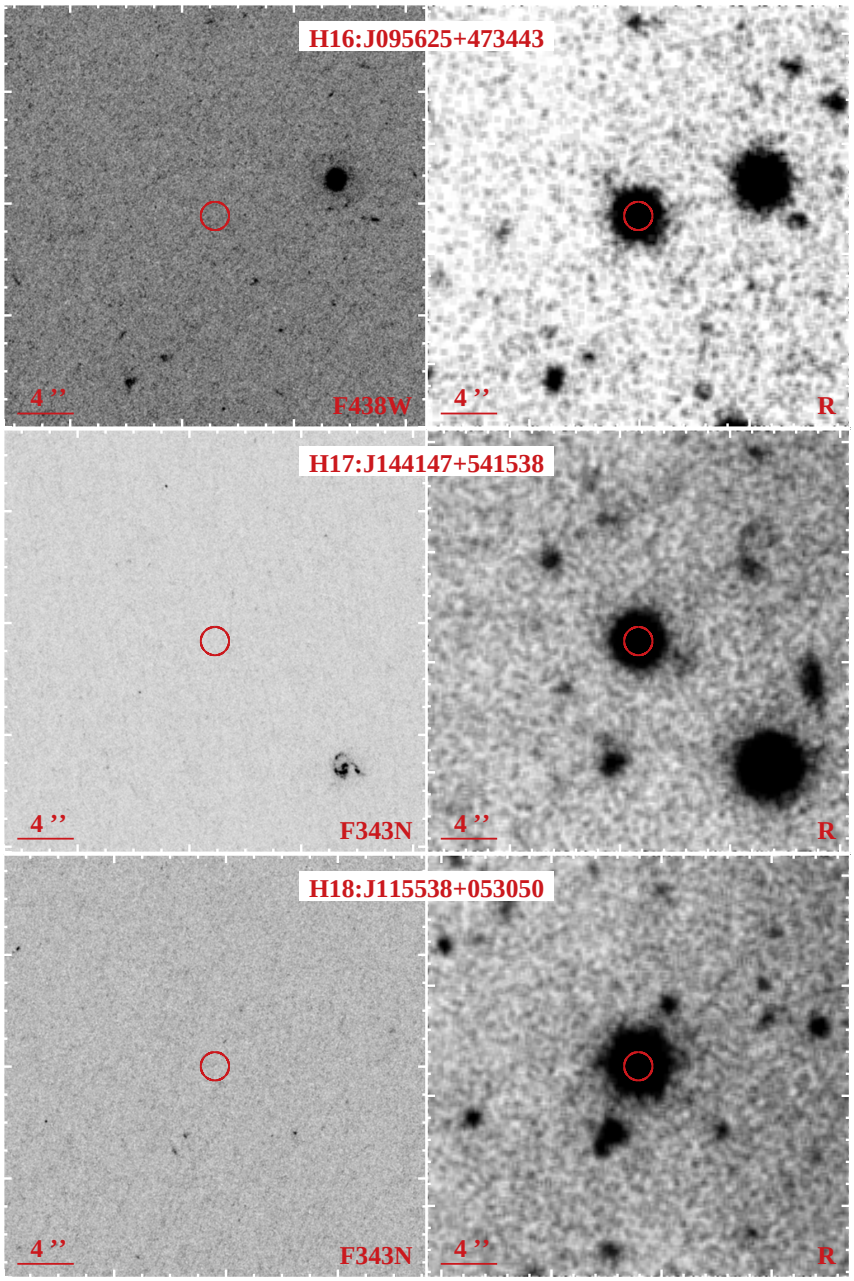

Figure 5.11: Gallery of the imaged quasar fields (continued).

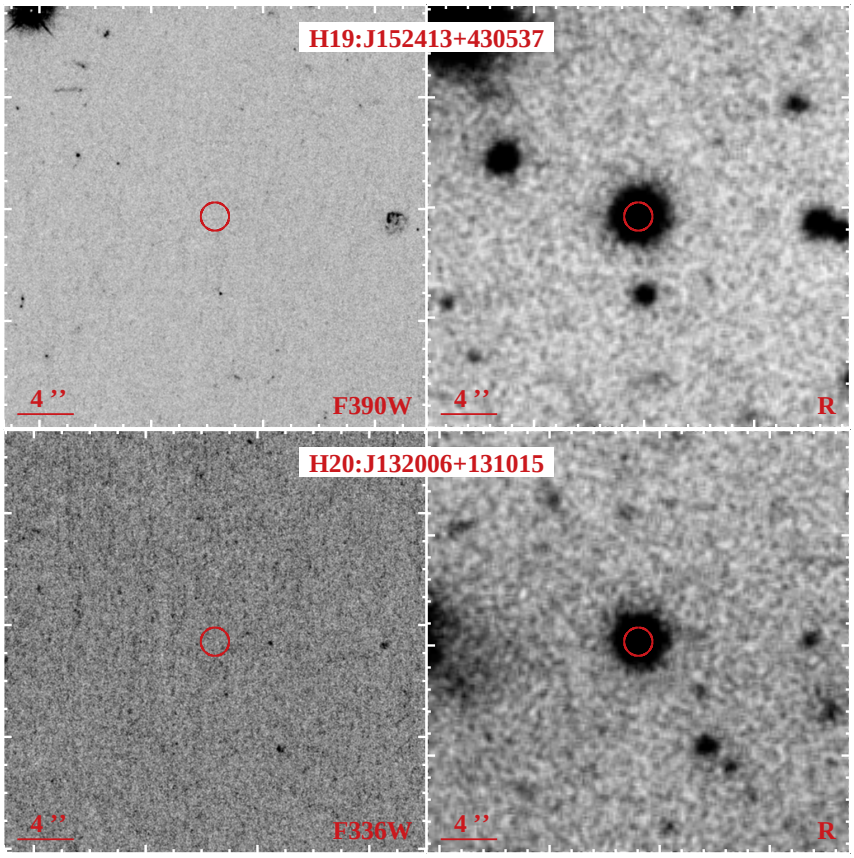

Figure 5.12: Gallery of the imaged quasar fields (continued).

and $I-$ band filters to the $r_{\rm sdss}$ and $i_{\rm sdss}$, respectively. To increase the number of sources that can be used for the photometric calibration, we group sets of images that were acquired with the same instrument and in similar weather conditions. A color term and an airmass correction are also included in our photometric solution. The residuals on the calibrated fluxes compared to the SDSS photometry are distributed around zero with a dispersion of ~ 0.1 mag (and some outliers at ~ 0.15 mag). For this reason, we believe that our photometry is accurate to within $\sim 10 - 15\%$.

Next we measure the minimum flux that can be detected in our imaging as a function of different aperture sizes, by modeling the noise across each image. This is done taking into account that the algorithms adopted during the coaddition of individual frames introduce a correlation in the noise of adjacent pixels. Specifically, we measure the standard deviation of a distribution of pixels extracted from apertures of progressively increasing sizes and that do not overlap with any of the detected sources. By varying the aperture size, the noise varies proportionally to the number of pixels as $N_{\rm pix}^{\beta}$. The index β describes the degree of correlations among pixels, and ranges from $\beta = 0.5$ in the case of uncorrelated noise to $\beta = 1$ in the opposite case. For each coadded image, we fit a model for the effective noise of the form $\sigma(N_{\rm pix}) = \sigma(1)\alpha N_{\rm pix}^{\beta}$, where $\sigma(1)$ is the noise per pixel and α is a coefficient of order of unity that accounts for small normalization corrections. Typical values of β in our imaging range from 0.6 to 0.75. Using this model, we characterize the depth of each image in circular apertures of $1''$ in diameter, as listed in Table 5.1.

Table 5.2: Log book of the quasar spectroscopic observations and summary of the DLA absorption properties.

Field	UT Date	Instr.	Slit ($''$)	Exp. Time (s)	$z_{\rm qso}$	$z_{\rm dla}$	$z_{\rm blo}$	$N_{\rm HI,dla}$ (log cm^{-2})	[X/H]$_{\rm dla}$	Ion
1:K1	2010 Jan, 5$^{\rm th}$	ESI	0.5	1800	3.424	2.9181	3.4420	20.30	-0.85 ± 0.12	Si
2:K2	2010 Jan, 5$^{\rm th}$	ESI	0.5	1800	3.676	2.6878	3.5515	20.65	-1.50 ± 0.17	Si
3:K3	2010 Jan, 5$^{\rm th}$	ESI	0.5	2400	3.427	2.3887	3.3958	21.05	-0.95 ± 0.17	Zn
4:K4	2010 Jan, 5$^{\rm th}$	ESI	0.5	3600	3.616	2.6880	3.3652	20.65	-1.32 ± 0.20	Fe
5:K5	2012 Jan, 25$^{\rm th}$	ESI	0.75	3600	3.623	2.5713	3.5071	20.75	-1.45 ± 0.21	Fe
6:K6	2010 Jan, 5$^{\rm th}$	ESI	0.5	3600	4.694	3.7861	4.5835	20.90	-2.08 ± 0.17	Fe
7:K7	2010 Jan, 5$^{\rm th}$	ESI	0.5	1800	3.880	2.7544	3.6287	20.20	-0.19 ± 0.10	Si
8:K9	2012 Jan, 25$^{\rm th}$	ESI	0.75	2400	3.824	2.7584	3.4559	20.90	-1.00 ± 0.10	Si
9:K10	2010 Jan, 5$^{\rm th}$	ESI	0.5	1800	3.629	2.4592	3.4812	20.35	-0.57 ± 0.10	Si
10:K11	2010 Jan, 5$^{\rm th}$	ESI	0.5	3600	4.499	3.5297	4.4671	21.10	-1.00 ± 0.16	Zn
11:K12	2010 Jan, 5$^{\rm th}$	ESI	0.5	1800	3.435	2.6606	3.4759	20.80	-1.41 ± 0.12	Si
12:K13	2010 Jan, 5$^{\rm th}$	ESI	0.5	2400	3.581	2.5978	3.4193	20.95	-1.33 ± 0.11	Si
13:H1	2010 Jan, 5$^{\rm th}$	ESI	0.5	1800	3.583	2.7803	3.6251	20.80	-1.68 ± 0.15	Fe
14:H2	VLT Archive	UVES	-	-	3.000	1.9127	2.6215	20.25	-	-
15:H3	2010 Jan, 5$^{\rm th}$	ESI	0.5	2400	3.996	3.2530	3.9147	20.65	-0.88 ± 0.11	Si
16:H4	2010 Jan, 5$^{\rm th}$	ESI	0.5	2400	3.582	2.7067	3.4366	20.40	-1.90 ± 0.16	Fe
17:H5	-	SDSS	-	-	3.719	3.0010	3.6373	21.10	-	-
18:H6	-	SDSS	-	-	3.710	2.9594	3.4071	20.95	-	-
19:H7	-	SDSS	-	-	4.133	3.3069	4.1215	20.55	-	-
20:H8	-	SDSS	-	-	3.175	2.6320	3.1124	20.40	-	-
21:H9	-	SDSS	-	-	2.482	1.8639	2.4762	20.95	-	-
22:H10	2010 Jan, 5$^{\rm th}$	ESI	0.5	2400	3.341	2.6826	3.2554	20.55	-1.21 ± 0.13	Si
23:H11	2012 Jan, 25$^{\rm th}$	ESI	0.75	3600	4.177	3.2332	3.8399	21.20	-1.26 ± 0.15	Fe
24:H12	-	SDSS	-	-	4.179	3.5645	4.1610	20.85	-	-
25:H13	2012 Jan, 25$^{\rm th}$	ESI	0.75	3600	3.073	2.6500	3.0427	20.70	-1.23 ± 0.15	Fe
26:H14	-	SDSS	-	-	3.804	2.7333	3.5865	20.35	-	-
27:H15	-	SDSS	-	-	3.998	2.9236	3.3930	20.55	-	-
28:H16	2001 Jan, 22$^{\rm th}$	ESI	0.5	4800	4.478	3.4035	4.2441	21.15	-1.75 ± 0.29	Fe

Continued on Next Page...

Table 5.2 – Continued

Field	UT Date	Instr.	Slit ($''$)	Exp. Time (s)	$z_{\rm qso}$	$z_{\rm dla}$	$z_{\rm blo}$	$N_{\rm HI,dla}$ (log cm^{-2})	[X/H]$_{\rm dla}$	Ion
29:H17	-	SDSS	-	-	3.467	2.6289	3.3336	20.85	-	-
30:H18	2006 Mar, 5$^{\rm th}$	ESI	0.75	3600	3.475	2.6079	3.3260	20.40	-2.23 ± 0.20	Si
31:H19	-	SDSS	-	-	3.920	2.8721	3.8787	20.45	-	-
32:H20	-	SDSS	-	-	3.352	2.6731	3.3417	20.35	-	-

5.4.2 Spectroscopic observations

We retrieved from the SDSS the spectra for all but one quasar (H2) which we use to establish the DLA redshifts and assess the hydrogen column density. To obtain an improved estimate of the hydrogen column density and metallicity of these DLAs, we also obtained Keck/ESI echellette spectra of 18 QSOs in our sample. Observations were conducted in good weather and seeing conditions using a $0.5''$ slit that yields a velocity resolution of ~ 37 km s^{-1}. Some of the observations were instead conducted with a $0.75''$ slit, corresponding to a velocity resolution of ~ 56 km s^{-1}. A summary of the spectroscopic observations and the corresponding exposure times is provided in Table 5.2. Data were reduced using the `ESIRedux` package[2] that processes the 2D images (bias subtraction, flat field), creates a wavelength solution and extracts, fluxes, and coadds the 1D spectra. In addition to these newly acquired spectra, we add to our spectroscopic sample three other quasars that are available through the Keck and VLT archives.

5.4.3 Column density and metallicity estimates

We establish precise redshifts for the DLAs and the high-redshift blocking systems (Table 5.2) using metal absorption lines that are free from contamination due to the forest. Whenever possible, we use low ionization lines that should provide us with a more robust estimate of the systemic redshift. After establishing the redshift for the main hydrogen components, we measure the hydrogen column density for the DLAs by modeling the absorption line with a Voigt profile. Since the line wings are sensitive to the hydrogen column density, we can determine the $N_{\rm HI}$ in DLAs to within $\sim 0.1 - 0.15$ dex, the major source of uncertainty being the placement of the intrinsic QSO continuum. Continuum estimates are in fact affected both by the S/N of each spectrum and by the absorption in the forest. For the SDSS spectra at lower resolution and lower S/N, the uncertainty on the final H I column density is larger, ~ 0.2 dex.

[2] www2.keck.hawaii.edu/inst/esi/ESIRedux

Figure 5.13 shows the modeled Voigt profiles and the associated uncertainties for each DLA.

For the 18 DLAs with new ESI spectra, we measure (or place upper limits on) the column density of ions in different ionization states with the apparent optical depth method (Savage & Sembach 1991). We examine each transition to select the ideal velocity window that would encompass the full line profile without including additional unrelated absorption. We also reject metal lines that are clearly contaminated by additional absorption from other hydrogen or metal lines, especially in the forest. Finally, we consider as lower limits the column density for the saturated lines. Tables for the metal column densities in these new 18 DLAs are provided in the Appendix 5.7.

Since the DLA gas is neutral, we do not have to apply ionization corrections that are instead essential at lower column densities (e.g. Fumagalli et al. 2011b). The gas phase metallicity can instead be obtained by comparing directly the hydrogen column density and the metal column density of elements that traces the bulk of the metal mass in DLAs (for a detailed discussion, see Rafelski et al. 2012). Ideally, one would adopt O or C, i.e. the most abundant elements, but unfortunately the only transitions of these elements that can be easily detected in DLAs are often saturated. In most of the cases, we rely on other α elements, and in particular on Si which has the convenient characteristic of exhibiting several transitions with different oscillator strengths in the UV. On the down side, Si is a refractory element and its gas phase abundance may underestimate the underlying metal content of the system. If Si transitions cannot be used, we measure the metallicity with Zn, although this element traces only a negligible fraction of the metal mass. Finally, if none of these elements is available, we measure the metallicity using Fe and applying an empirical correction of $[X/H] = [Fe/H] + 0.3$ dex (see Rafelski et al. 2012). In this study, we assume solar abundances from Asplund et al. (2009) and atomic data from Morton (2003). A summary of the derived metallicities is in Table 5.2.

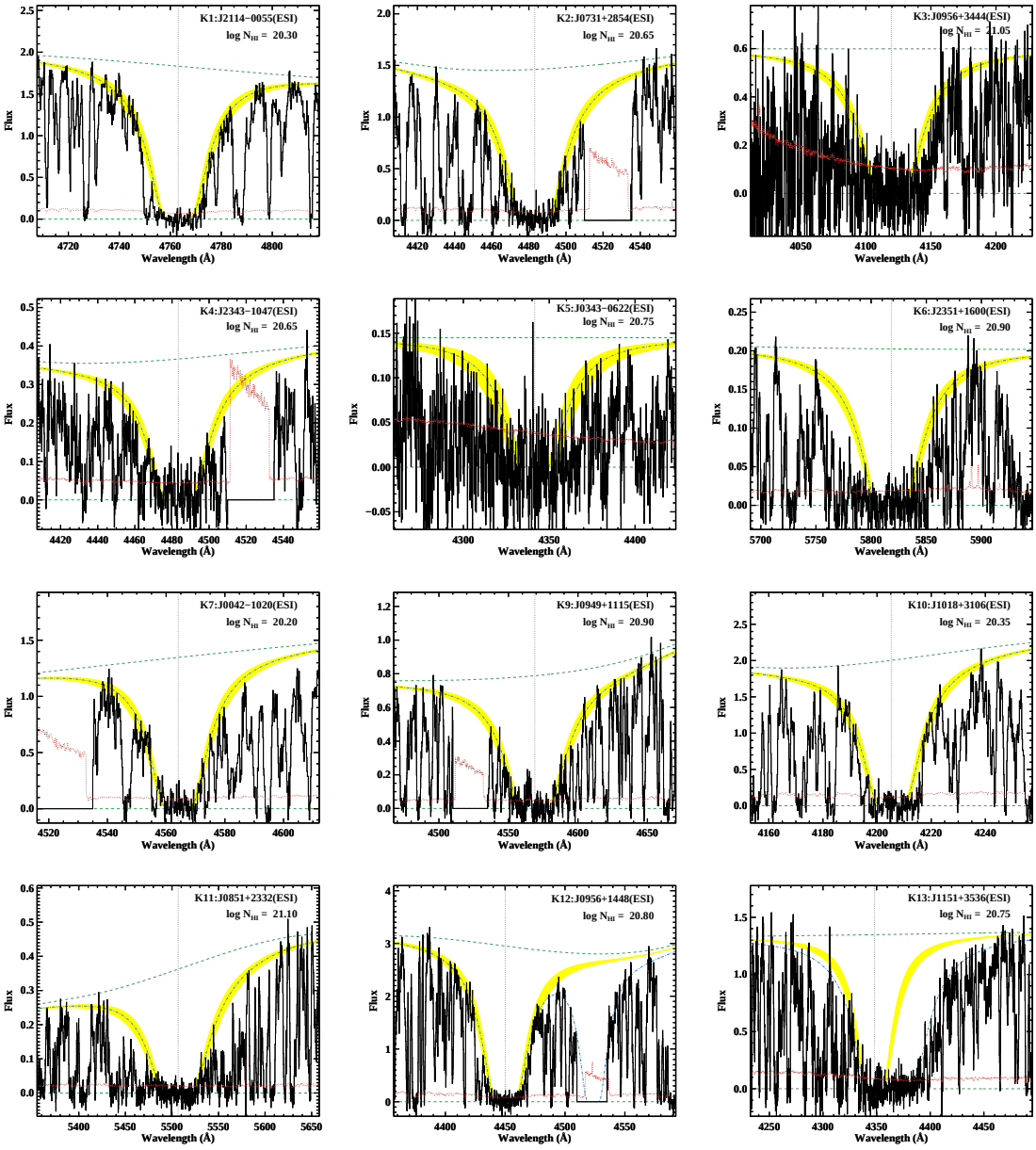

Figure 5.13: Voigt profiles of the Lyα absorption lines. In each panel, we superimpose to the data (black histograms) the quasar continuum level (green dashed line), the absorption line models (blue dash-dotted line), and the corresponding 1σ errors for the main hydrogen component only (yellow shaded regions). Uncertainties on the flux are shown by a red dotted line, while the systemic redshift of each DLA is marked by vertical gray dotted line.

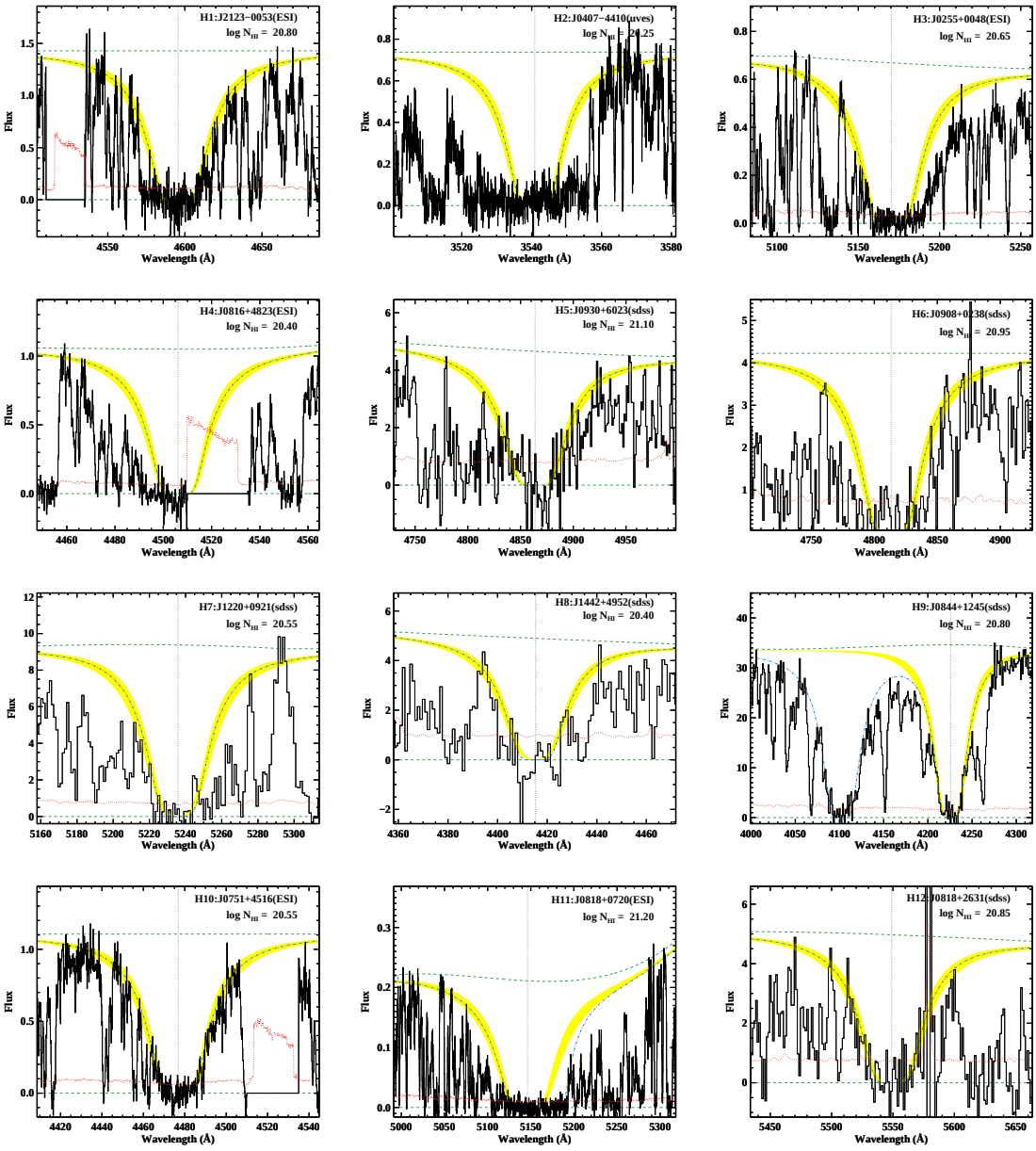

Figure 5.14: Voigt models of the Lyα absorption lines (continued).

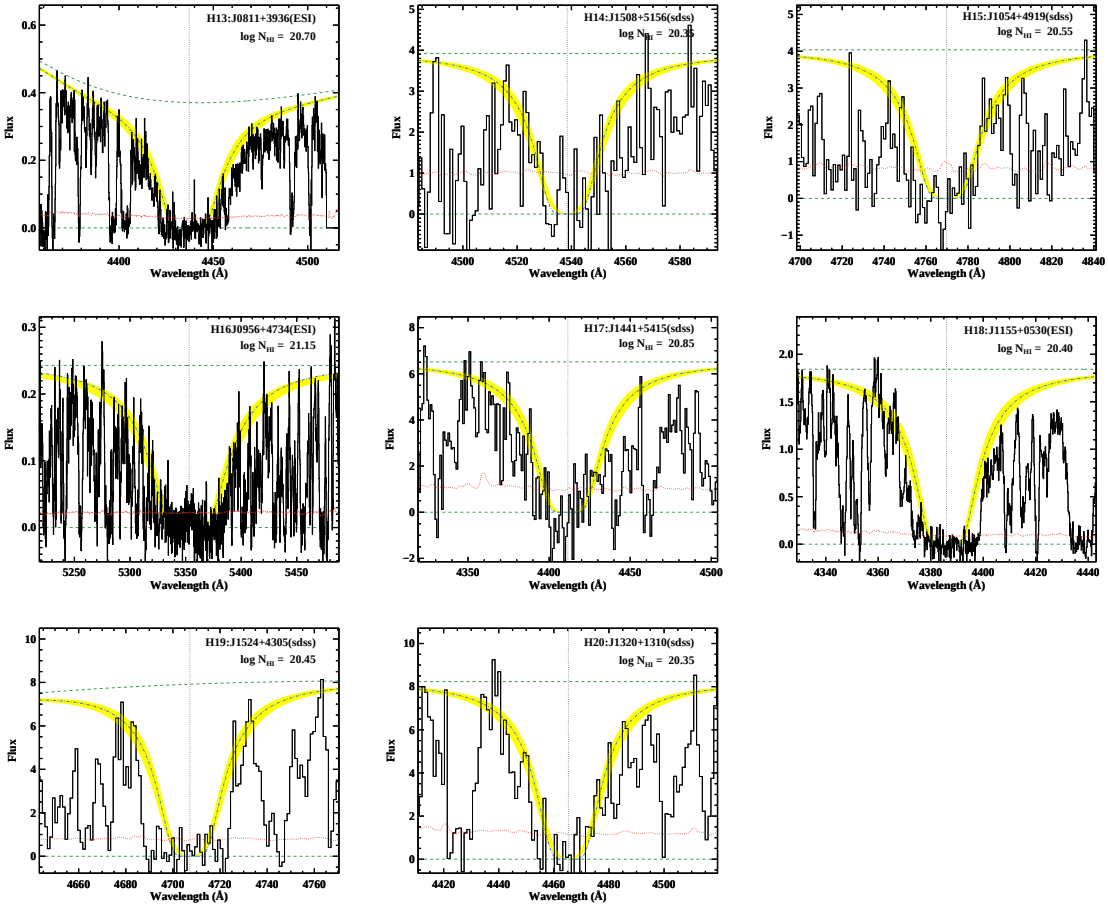

Figure 5.15: Voigt models of the Lyα absorption lines (continued).

5.4.4 Star formation rate conversions

To convert the observed fluxes to SFRs, we adopt the following conversions (see Paper I for a detailed discussion):

$$\frac{SFR}{\mathrm{M}_{\odot}\mathrm{yr}^{-1}} = 7.91 \times 10^{-29} \times f_{\nu} \times 4\pi D^2 \times K_{\mathrm{corr}} \times \phi_{\mathrm{igm}} \times \chi_{\mathrm{MW}} \,. \tag{5.1}$$

In this equation, the constant 7.91×10^{-29} converts the rest-frame FUV luminosity to a SFR (Madau et al. 1998), D is the luminosity distance to the DLA, K_{corr} is the K-correction factor for a flat spectral energy distribution (SED), ϕ_{igm} accounts for the absorption due to the intervening intergalactic medium for photons bluer than Lyα, and χ_{MW} accounts for the Milky Way extinction. We neglect intrinsic dust absorption. In deriving the flux f_{ν}, we apply the appropriate photometric calibration and, where measured colors are not available, we assume a typical color for a star forming galaxy at the corresponding redshift of each DLA.

5.5 Analysis and discussion

Figure 5.2 reveals that the main strength of this imaging survey is the possibility to search for faint ($< 1\mathrm{M}_{\odot}$ yr^{-1}) star forming galaxies at any projected distance from the background QSOs. However, lacking spectroscopic follow-up of the detected galaxies, we cannot provide a unique associations between the DLAs and individual galaxies. Yet, we can take advantage of the large (for these types of studies!) statistics to infer the characteristic *in situ* SFR for DLA galaxies. We can further study the environment around these lines of sight to probe the connection between DLAs and optically selected galaxies, again on a statistical ground.

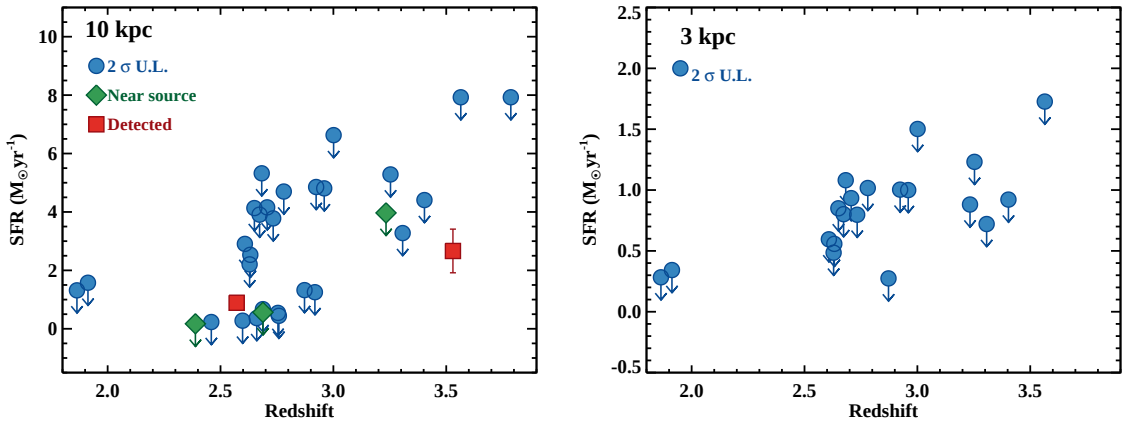

Figure 5.16: Local SFRs in the observed DLAs. On the left, we show 2σ upper limits (blue circles) and detections (at 3σ CL; red squares) computed for a circlar aperture of 10 kpc centered at the quasar position. DLAs for which a nearby galaxy leaks some light within the aperture without resulting in a 3σ detection are marked with green diamonds. On the right, we show 2σ limits computed in a smaller aperture of 3 kpc for the HST sample only. Regardless to whether star formation occurs in compact or extended regions, we infer that the majority of DLAs have an *in situ* SFR of $\lesssim 1$ $M_{\odot}$ yr^{-1}

5.5.1 The *in situ* SFR of DLAs

Turning our attention to the problem of characterizing the *in situ* SFR, we search for emission in the $U-$ and $B-$band images at the projected location of the QSOs. First, we consider apertures of 10 kpc, which plausibly encompass all the light from an extended gaseous disk at high redshift. At the median DLA redshift of our sample, an aperture of 10 kpc corresponds to $\sim 1.2''$. Thus, for this experiment, we can use both HST and ground based imaging. The measured SFRs are shown in the left panel of Figure 5.16. These are computed assuming that all the emission originates from the DLA redshift.

Emission at the projected QSO position is detected at 3σ C.L. in 2/32 cases ($< 10\%$), along the line of sight to K5 and K11. Inspecting the imaging, we see that the emission from K5 is slightly off center, while the emission from K11 is aligned with the projected position of the background quasar. Lacking spectroscopic observations, we can not confirm that the detected emission is in fact associated to the DLAs. However, the probability that at least one random

interloper lies within $\sim 1''$ from the projected QSO position is $< 10\%$ for galaxies of ~ 27 mag (see Paper I). Further, from the Voigt profile of the Lyα line, we infer a column density of $\sim 10^{20}$ cm^{-2} for the high-redshift blocking system in the K5 sightline. This corresponds to an opacity at the Lyman limit of ~ 600 and thus the detected flux cannot be from the background quasar. Similarly, the transmitted quasar flux for K11 is completely suppressed by a strong proximity Lyman limit system whose Lyman series is saturated in the ESI spectrum to at least Ly13. Finally, none of the two lines of sight shows a strong absorber ($N_{\rm HI} \gtrsim 10^{19}$cm^{-2}) in between the targeted DLA and the blocking filter, reducing the chance that we are detecting the host galaxy of another absorption system that is unrelated to the DLA. For all these reasons, we regard the detected emission at the project QSO position as coming from the DLA galaxy counterparts. At the corresponding redshift, we infer SFRs of $0.9 M_\odot$ yr^{-1} and $2.7 M_\odot$ yr^{-1} for K5 and K11, respectively.

In Figure 5.16, we also show 2σ upper limits of the remaining 30 sightlines with no detected emission at the projected QSO location. We also highlight three DLAs (K3, K4, and H11) for which some flux from a galaxy in proximity to the aperture is detected within 10 kpc, but for which the total flux within the aperture does not exceed the 2σ CL. Upper limits on the SFRs range from $\sim 0.2 M_\odot$ yr^{-1} to $\sim 8 M_\odot$ yr^{-1}, with a mean of ~ 3 $M_\odot$ yr^{-1}. The ground-based imaging provides the most stringent constraints for extended emission ($\sim 1 M_\odot$ yr^{-1}). Conversely, due to the superior resolution, the HST imaging allows us to probe the *in situ* SFR from compact objects. In fact, by considering an aperture of 3 kpc at the DLA redshift, we place upper limits (shown in the right panel of Figure 5.16) to the local SFRs of compact sources that range between 0.3 $M_\odot$ yr^{-1} to 1.7 $M_\odot$ yr^{-1}, with a mean of $\sim 0.8 M_\odot$ yr^{-1}. Combined together, these observations constrain the *in situ* SFR of a typical DLA to be less than $\sim 1 M_\odot$ yr^{-1}either for an extended low surface brightness object or for a compact galaxy. Recast in terms of the star formation surface density $\Sigma_{\rm sfr}$, typical DLAs form stars to less than $< 0.01 - 0.03$ $M_\odot$ yr^{-1}

kpc^{-2}.

The lack of *in situ* star formation to faint levels is consistent with the results of Wolfe & Chen (2006). These authors placed even tighter ($\times 2 - 2.5$) limits to the star formation surface density in the Hubble ultra deep field by searching for extended emission associated to a putative population of galaxies whose number density would match the frequency with which DLAs are detected. Our analysis not only confirms the findings of Wolfe & Chen (2006), but it reinforces this previous result in two ways. First, the study Wolfe & Chen (2006) did not consider compact sources, which are included in our measurements. Second, we are directly searching for emission at the position of a known DLA rather than connecting the emission and absorption with a statistical formalism.

Tight limits on the local specific star formation are also expected on theoretical grounds. In recent years, a large number of theoretical and observational studies have addressed the problem of the formation of stars in regions that are dominated by atomic hydrogen (e.g. Krumholz et al. 2008; 2009a, Fumagalli et al. 2009; 2010a, Bigiel et al. 2010, Bolatto et al. 2011, Glover & Clark 2012). The general consensus is that the ability of the gas to form stars depends on the cooling properties which, in turn, are sensitive to the amount of dust and gas column density that can shield the gas from the ambient radiation field. As a consequence, star formation proceeds inefficiently at low hydrogen column densities and in metal poor environments.

In Figure 5.17, we present our estimates for the *in situ* star formation surface density as a function of the H I column density measured in absorption. Superimposed to the data, we show the original determination of the Kennicutt-Schmidt law from Kennicutt (1998) and two sets of models for the star formation law from Krumholz et al. (2009b) at solar and subsolar metallicity. This comparison reveals that, according to models, the hydrogen column density and the low metal content in a typical DLA are insufficient to sustain appreciable star formation (see Krumholz et al. 2009c). It must be noted that the inferred star formation surface density is

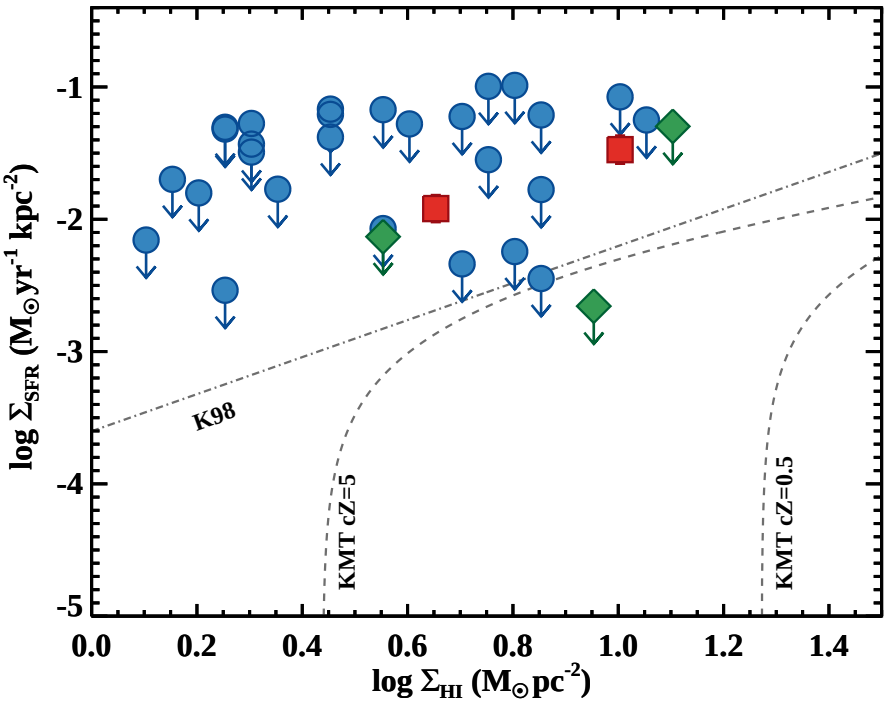

Figure 5.17: The $\Sigma_{\rm SFR}/\Sigma_{\rm HI}$ relation for the imaged DLAs. Since the star formation surface density is computed within an aperture of 10 kpc while the H I column density is probed along the line of sight to the quasars, this relation is not strictly equivalent to the Kennicutt-Schmidt law. Also plotted, the original KS law (Kennicutt 1998) and models from Krumholz et al. (2009b) for different metallicities and clumpiness factors (dashed lines). The local gas column density in these DLAs is not enough to sustain appreciable *in situ* star formation.

measured on scales of > 1 kpc, while the H I column density is derived from a pencil beam on parsec scales. It is therefore unclear whether the H I column density reflects the average value on larger scales. Further, because of the low covering factor of molecular gas and the high dust extinction in molecular regions, we do not expect that molecular gas contributes significantly to the total gas column density along the line of sight (Krumholz et al. 2009c). Only the two DLAs that are detected in emission appear to form star at even higher rate than what is expected from the star formation laws (see a similar discussion in Wolfe & Chen 2006). In this case, however, we cannot exclude the presence of significant molecular gas that is displaced from the line of sight to the quasar.

5.5.2 DLAs and LBGs

The connection between DLAs and LBGs has been often explored in the literature (e.g. Fynbo et al. 1999, Møller et al. 2002; 2004, Wolfe & Chen 2006, Christensen et al. 2007, Fumagalli et al. 2010b, Rafelski et al. 2011). By definition, LBGs are star-forming galaxies that are selected based on colors at optical wavelengths searching for the characteristic absorption in the UV photons blueward to Lyα. For this reason, LBGs are subject to optical selection effects and surveys that reach different depths do not necessarily probe identical populations. To avoid this confusion, we recast this problem in different terms, i.e. focusing on the relationship between DLAs and star-forming galaxies. Specifically, we wish to test the three scenarios outlined in the introduction: do DLAs arise from the outskirt of bright star forming galaxies, or from faint dwarf galaxies in isolation, or again from clumps and satellites within the halo of a more massive central galaxy?

A detailed study of the environment around the targeted DLAs requires extensive spectroscopic follow-up. However, we can still constrain some of the above scenarios by performing the following experiment. In Figure 5.18, we show the distribution of impact parameters to the

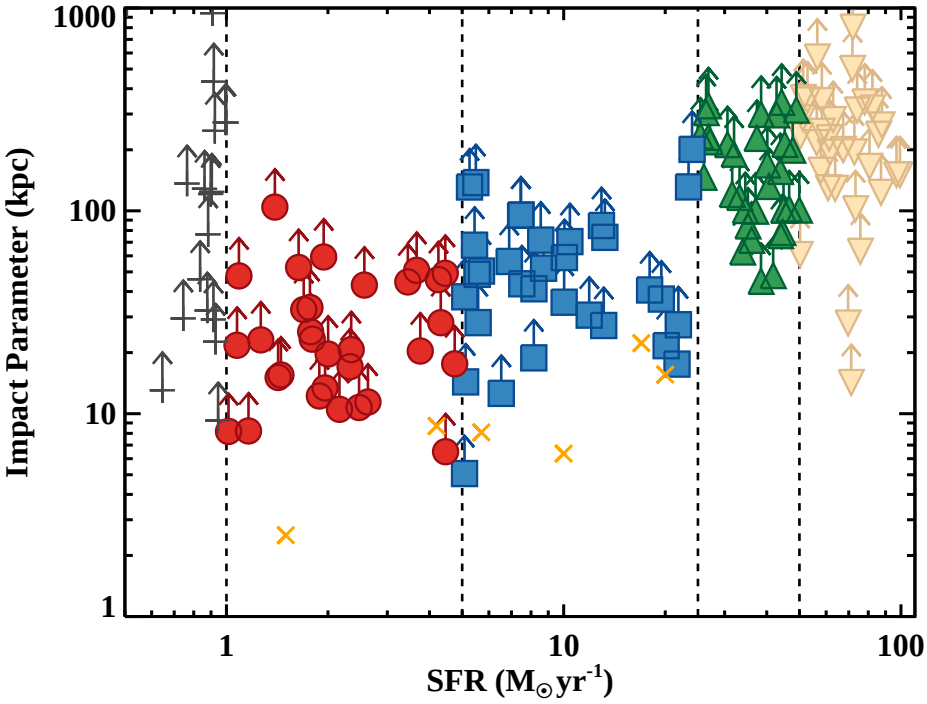

Figure 5.18: Distribution of the impact parameters to the closest detected galaxies in four bins of SFRs ($1-5M_\odot$ yr^{-1} red circles; $5-25M_\odot$ yr^{-1} blue squares; $25-50M_\odot$ yr^{-1} green upward triangles; $50-100M_\odot$ yr^{-1} tan downward triangles). We also consider galaxies with SFR$< 1M_\odot$ yr^{-1}(gray crosses) to show that we are not missing a large number of galaxies with SFR$\sim 1M_\odot$ yr^{-1}at close impact parameters. Also plotted, six DLAs with known emitting counterparts from the literature (orange Xs). Since all the galaxies are assumed to be at the DLA redshifts, these impact parameters are formally lower limits. $\sim 90\%$ of the detected galaxies lie beyond 10 kpc.

closest galaxy within a given interval of SFRs. This is computed by converting the observed fluxes of the galaxies detected in each field to SFRs at the DLA redshift and by measuring the distance from the projected quasar position to the *closest* detected galaxy in four intervals of SFRs ($1-5 M_\odot$ yr^{-1}, $5-25 M_\odot$ yr^{-1}, $25-50 M_\odot$ yr^{-1}, and $50-100 M_\odot$ yr^{-1}). Since these galaxies may be interlopers, the quoted impact parameters are formally lower limits. We limit our search to SFR$> 1 M_\odot$ yr^{-1} since Figure 5.16 shows that we are almost complete down to this level. Despite the higher degree of incompleteness, we also show the distribution of impact parameters for galaxies with SFR$< 1 M_\odot$ yr^{-1} to verify that our results are not particularly sensitive to the imposed cut in SFR.

In the SFR interval $1-25$ $M_\odot$ yr^{-1}, none of the detected galaxies lie within 5 kpc from the projected QSO position (except for K11 where we detect *in situ* SFR from an object centered at the QSO location). In the SFR interval $1-5 M_\odot$ yr^{-1}, the fraction of DLAs for which the closest galaxy lie outside 10 kpc, 20 kpc and 30 kpc is respectively $> 90\%$, $> 59\%$, and $> 36\%$. These values do not appear to drastically change for $< 1 M_\odot$ yr^{-1}, although our completeness decreases significantly in this bin. For the SFR interval $5-25 M_\odot$ yr^{-1}, instead, the corresponding fractions are $> 97\%$, 84%, and 72%. Finally, $> 75\%$ ($> 87\%$) of the DLAs have the closest galaxy with SFR between $25-50 M_\odot$ yr^{-1} ($50-100 M_\odot$ yr^{-1}) beyond 100 kpc. Note that the probability to find at least one interloper brighter than an L_* galaxy at $z \sim 2.5$ within 50 kpc from the quasar is $\sim 20\%$ (see Paper I).

From this simple experiment, we infer that a typical DLA does not reside within the gaseous disk of a galaxy with SFR$> 1 M_\odot$ yr^{-1} (Prochaska & Wolfe 2009), unless the H I extends much beyond $10-20$ kpc. Conversely, our data do not rule out the possibility that DLAs arise from satellites within the virial radius of galaxies with SFR$> 1 M_\odot$ yr^{-1}. This means that either DLAs originates from very faint galaxies with SFR$< 1 M_\odot$ yr^{-1} that are centrals to their halos (i.e. $\lesssim 0.1 L_*$ Rauch et al. 2008) or from the surroundings of more massive galaxies $\gtrsim 0.1 L_*$ (see

Fumagalli et al. 2011a), but not much above L_* as indicated by the small fraction of galaxies with SFR$> 25 M_\odot$ yr^{-1} within 100 kpc from the quasar. Using a stack of 48 LBGs from the Hubble Ultra Deep Field, Rafelski et al. (2011) have found that the low level of star formation in the outskirts of LBGs ($\lesssim$ 10kpc) may be physically connected to DLAs. This result appears in tension with our findings, but the typical SFRs of their galaxies ranges from 0.5 to 6 $M_\odot$ yr^{-1}, with a median of $\sim 1 M_\odot$ yr^{-1}. These values are comparable to the detection limit of our imaging. Further, it should be noted that Rafelski et al. (2011) considered DLAs with higher column density ($\log N_{\rm HI} > 21$) than the ones in our sample. Thus, the two results may still be consistent.

Finally, we can compare our statistical sample to the properties of the six known counterparts of DLAs with published SFRs in the literature (table 1 in Krogager et al. 2012). Although the limited number of objects prevents a robust comparison, we see from Figure 5.18 that confirmed DLA counterparts are distributed at higher SFRs ($\gtrsim 5 M_\odot$ yr^{-1}) and at lower impact parameters than what is typical in our sample (4/6 DLA galaxies lie within 10 kpc). Although one may conclude that DLA galaxies are forming stars at more than few $M_\odot$ yr^{-1} and may arise from the surrounding of star forming galaxies, the comparison with our statistical sample suggests that the confirmed DLA galaxies form a biased sample that may not be representative of the typical DLA population. This is especially true given that the number of non-detections in previous searches for DLA counterparts is unknown and that often only the most metal rich DLAs were targeted.

5.6 Summary and conclusions

We have presented results from an imaging survey of 32 quasar fields with known damped Lyman-α (DLA) systems at $z \sim 2 - 3.5$. To avoid the quasar contamination that

hampered previous imaging studies of DLA galaxies, we have selected only sightlines in which optically thick absorbers at higher redshift fully block the quasar light at the corresponding rest-frame FUV frequencies of the DLA galaxies (O'Meara et al. 2006, Fumagalli et al. 2010b). In this way, we can search for faint star-forming galaxies at any impact parameter.

Our principal findings are here summarized.

- For 2/32 DLA fields we detect emission at the projected quasar position. Based on the low incidence of interlopers at comparable angular separations ($< 10\%$ for galaxies of ~ 27 mag) and on the fact that the high-redshift blocking filters are highly optically thick ($\tau \gg 2$), we consider the detected flux associated to the DLAs. At the corresponding redshift, we infer *in situ* SFRs for these two DLA galaxies of $0.9 M_\odot$ yr^{-1}and $2.7 M_\odot$ yr^{-1}.

- For the remaining 30 sightlines, we place upper limits to the *in situ* SFR of DLA galaxies at $\sim 1 M_\odot$ yr^{-1} both for extended (10 kpc) and compact (3 kpc) sources. These limits translate to star formation surface densities of $< 0.01 - 0.03$ $M_\odot$ yr^{-1} kpc^{-2}.

- By studying the environment around these sightlines, we find that $> 90\%$ ($> 97\%$) of the galaxies with SFRs between $1 - 5 M_\odot$ yr^{-1}($5 - 25 M_\odot$ yr^{-1}) lie outside 10 kpc from the projected quasar position. Further, the occurrence of $\gg L_*$ galaxies in the inner 100 kpc is rare.

- By comparing our statistical sample to six DLAs with known galaxy counterparts we find systematic differences among the two data sets, suggesting that perhaps the sample of confirmed DLA galaxies is not fully representative of the typical DLA population.

Based on these findings, we conclude that significant SFR is not occurring in the DLA gas. This result confirms the findings of Wolfe & Chen (2006) both for extended and compact sources. The lack of *in situ* SFR in DLAs is expected on theoretical grounds because the typical

hydrogen column density and low metal content prevent shielding and cooling of the gas. In turn, this inhibits significant star formation.

The study of the environment around DLAs allows us to further conclude that DLAs arise from very faint $\lesssim 0.1L_*$ galaxies that are centrals to their halos or from clumps and satellites in the halo of galaxies up to few L_*. Generally, we do not find evidence that typical DLAs originate in the outskirts (< 10kpc) of galaxies with SFR$> 1\mathrm{M}_\odot$ yr^{-1}.

5.7 Appendix

In this Appendix, we present the column densities of selected ionic transitions for 18 DLAs included in this study with new Keck/ESI echellette spectra.

Table 5.3: Table of the metal column densities for the targeted DLAs with newly acquired ESI spectra.

Ion	λ (Å)	$\log f$	Instr.	$v_{\rm int}$ (km s^{-1})	W_λ (mÅ)	$\log N$	$\log N_{\rm adopt}$
K1:J21144-0055 at $z = 2.91806$							
C I							
	1560.3092	−0.8808	ESI	$[-100, 100]$	< 18.0	< 12.98	< 12.98
	1656.9283	−0.8273	ESI	$[-100, 100]$	< 25.5	< 13.02	
C II							
	1334.5323	−0.8935	ESI	$[-50, 50]$	476.6 ± 3.3	> 14.85	> 14.85
C IV							
	1548.1950	−0.7194	ESI	$[-120, 120]$	482.5 ± 9.0	14.23 ± 0.01	14.23 ± 0.01
	1550.7700	−1.0213	ESI	$[-120, 100]$	290.3 ± 9.1	14.24 ± 0.02	
O I							
	1302.1685	−1.3110	ESI	$[-70, 90]$	380.8 ± 8.7	> 15.07	> 15.07
Al II							
	1670.7874	0.2742	ESI	$[-100, 100]$	504.7 ± 10.5	> 13.33	> 13.33
Al III							
	1854.7164	−0.2684	ESI	$[-150, 80]$	252.5 ± 11.2	> 13.31	> 13.31
	1862.7895	−0.5719	ESI	$[-100, 100]$	163.7 ± 13.9	< 13.38	
Si II							
	1190.4158	−0.6017	ESI	$[-120, 120]$	487.1 ± 14.4	> 14.51	14.96 ± 0.06
	1260.4221	0.0030	ESI	$[-100, 100]$	626.0 ± 10.5	> 14.03	
	1304.3702	−1.0269	ESI	$[-100, 100]$	336.2 ± 9.9	> 14.66	
	1526.7066	−0.8962	ESI	$[-120, 90]$	426.0 ± 10.3	> 14.55	
	1808.0130	−2.6603	ESI	$[-60, 60]$	52.7 ± 7.8	14.96 ± 0.06	
Si IV							
	1393.7550	−0.2774	ESI	$[-140, 100]$	351.6 ± 5.5	13.71 ± 0.01	13.75 ± 0.01
	1402.7700	−0.5817	ESI	$[-120, 90]$	260.8 ± 5.6	13.84 ± 0.01	

Continued on Next Page...

Table 5.3 – Continued

Ion	λ (Å)	$\log f$	Instr.	$v_{\rm int}$ (km s^{-1})	W_λ (mÅ)	$\log N$	$\log N_{\rm adopt}$
S II							
	1250.5840	−2.2634	ESI	$[-50, 50]$	< 20.3	< 14.63	< 14.63
	1253.8110	−1.9634	ESI	$[-50, 110]$	72.8 ± 12.2	< 14.73	
	1259.5190	−1.7894	ESI	$[-60, 60]$	115.2 ± 9.8	< 14.80	
Cr II							
	2056.2539	−0.9788	ESI	$[-70, 70]$	32.0 ± 10.3	12.93 ± 0.14	12.93 ± 0.14
	2062.2340	−1.1079	ESI	$[-70, 70]$	< 21.4	< 13.05	
	2066.1610	−1.2882	ESI	$[-70, 70]$	< 22.7	< 13.25	
Fe II							
	2249.8768	−2.7397	ESI	$[-60, 60]$	< 18.9	< 14.55	> 14.35
	2260.7805	−2.6126	ESI	$[-60, 60]$	< 19.4	< 14.43	
	2344.2140	−0.9431	ESI	$[-100, 100]$	488.0 ± 15.1	> 14.19	
	2374.4612	−1.5045	ESI	$[-80, 80]$	260.0 ± 20.2	> 14.35	
	2382.7650	−0.4949	ESI	$[-100, 100]$	819.0 ± 22.7	> 14.01	
Ni II							
	1454.8420	−1.4908	ESI	$[-50, 50]$	< 13.8	< 13.53	< 13.31
	1709.6042	−1.4895	ESI	$[-50, 50]$	30.9 ± 6.8	< 13.60	
	1741.5531	−1.3696	ESI	$[-50, 50]$	28.3 ± 4.4	< 13.41	
	1751.9157	−1.5575	ESI	$[-50, 50]$	< 10.2	< 13.31	
Zn II							
	2026.1360	−0.3107	ESI	$[-40, 40]$	< 14.5	< 12.10	< 12.10
K2:J0731+2854 at $z = 2.68778$							
C I							
	1560.3092	−0.8808	ESI	$[-100, 100]$	< 13.1	< 12.84	< 12.77
	1656.9283	−0.8273	ESI	$[-40, 50]$	< 14.0	< 12.77	
C II							

Continued on Next Page...

Table 5.3 – Continued

Ion	λ (Å)	$\log f$	Instr.	$v_{\rm int}$ (km s^{-1})	W_λ (mÅ)	$\log N$	$\log N_{\rm adopt}$
	1334.5323	−0.8935	ESI	$[-70, 70]$	304.9 ± 10.1	> 14.49	> 14.49
C IV							
	1548.1950	−0.7194	ESI	$[-100, 100]$	116.9 ± 5.6	13.50 ± 0.02	13.50 ± 0.02
	1550.7700	−1.0213	ESI	$[-100, 100]$	113.3 ± 5.7	< 13.81	
O I							
	1302.1685	−1.3110	ESI	$[-30, 30]$	290.8 ± 3.7	> 15.06	> 15.06
Al II							
	1670.7874	0.2742	ESI	$[-70, 70]$	210.7 ± 7.8	> 12.87	> 12.87
Al III							
	1854.7164	−0.2684	ESI	$[-70, 70]$	< 18.4	< 12.23	< 12.23
Si II							
	1193.2897	−0.3018	ESI	$[-70, 70]$	337.6 ± 11.7	> 14.08	14.66 ± 0.14
	1260.4221	0.0030	ESI	$[-80, 60]$	378.4 ± 11.7	> 13.75	
	1304.3702	−1.0269	ESI	$[-40, 40]$	243.4 ± 5.4	> 14.48	
	1808.0130	−2.6603	ESI	$[-50, 50]$	26.9 ± 8.6	14.66 ± 0.14	
Si IV							
	1393.7550	−0.2774	ESI	$[-120, 100]$	570.7 ± 9.8	< 14.04	< 13.82
	1402.7700	−0.5817	ESI	$[-100, 200]$	257.3 ± 13.0	< 13.82	
Mn II							
	2594.4990	−0.5670	ESI	$[-50, 50]$	< 57.2	< 12.72	< 12.72
	2606.4620	−0.7151	ESI	$[-50, 50]$	< 61.2	< 12.91	
Fe II							
	1608.4511	−1.2366	ESI	$[-50, 50]$	155.1 ± 7.1	> 14.23	14.59 ± 0.11
	1611.2005	−2.8665	ESI	$[-50, 50]$	< 15.8	< 14.89	
	2249.8768	−2.7397	ESI	$[-40, 40]$	< 22.7	< 14.64	
	2260.7805	−2.6126	ESI	$[-40, 40]$	40.6 ± 10.5	14.59 ± 0.11	
	2344.2140	−0.9431	ESI	$[-70, 60]$	350.7 ± 9.1	> 14.06	
	2374.4612	−1.5045	ESI	$[-50, 50]$	198.1 ± 7.9	> 14.23	
	2382.7650	−0.4949	ESI	$[-60, 60]$	396.3 ± 7.9	> 13.71	
	2586.6500	−1.1605	ESI	$[-50, 50]$	309.6 ± 25.7	> 14.04	

Continued on Next Page...

Table 5.3 – Continued

Ion	λ (Å)	$\log f$	Instr.	$v_{\rm int}$ (km s^{-1})	W_λ (mÅ)	$\log N$	$\log N_{\rm adopt}$
	2600.1729	−0.6216	ESI	$[-60, 60]$	417.3 ± 23.5	> 13.73	
Ni II							
	1709.6042	−1.4895	ESI	$[-40, 40]$	< 14.7	< 13.43	< 13.43
	1741.5531	−1.3696	ESI	$[-30, 80]$	89.3 ± 8.7	< 13.98	
	1751.9157	−1.5575	ESI	$[-50, 50]$	29.4 ± 8.4	< 13.64	
Zn II							
	2026.1360	−0.3107	ESI	$[-30, 30]$	< 13.8	< 12.07	< 12.07
				K3:J0956+3444 at $z = 2.38874$			
C I							
	1656.9283	−0.8273	ESI	$[-50, 50]$	< 20.3	< 12.91	< 12.91
C IV							
	1548.1950	−0.7194	ESI	$[-140, 150]$	1238.3 ± 11.7	> 14.90	> 14.90
	1550.7700	−1.0213	ESI	$[-140, 150]$	765.3 ± 13.8	> 14.78	
O I							
	1302.1685	−1.3110	ESI	$[-150, 150]$	886.4 ± 23.6	> 15.45	> 15.45
Al II							
	1670.7874	0.2742	ESI	$[-130, 130]$	778.6 ± 15.1	> 13.62	> 13.62
Al III							
	1854.7164	−0.2684	ESI	$[-80, 80]$	240.7 ± 14.1	13.25 ± 0.03	13.25 ± 0.03
Si II							
	1260.4221	0.0030	ESI	$[-120, 120]$	1021.9 ± 32.7	> 14.26	> 16.07
	1304.3702	−1.0269	ESI	$[-120, 120]$	780.9 ± 20.4	> 15.13	
	1526.7066	−0.8962	ESI	$[-180, 180]$	877.5 ± 15.9	> 14.89	
	1808.0130	−2.6603	ESI	$[-120, 120]$	444.6 ± 13.9	> 16.06	
Si IV							
	1393.7550	−0.2774	ESI	$[-120, 80]$	614.5 ± 16.6	> 14.10	> 14.10

Continued on Next Page...

Table 5.3 – Continued

Ion	λ (Å)	$\log f$	Instr.	$v_{\rm int}$ (km s^{-1})	W_λ (mÅ)	$\log N$	$\log N_{\rm adopt}$
Cr II							
	2056.2539	-0.9788	ESI	$[-120, 120]$	129.5 ± 17.7	13.56 ± 0.06	13.57 ± 0.05
	2062.2340	-1.1079	ESI	$[-120, 120]$	130.6 ± 18.0	< 13.69	
	2066.1610	-1.2882	ESI	$[-120, 120]$	72.7 ± 18.4	13.62 ± 0.10	
Mn II							
	2594.4990	-0.5670	ESI	$[-60, 60]$	116.5 ± 18.6	12.91 ± 0.07	12.91 ± 0.07
	2606.4620	-0.7151	ESI	$[-100, 100]$	224.2 ± 42.7	< 13.37	
Fe II							
	1608.4511	-1.2366	ESI	$[-150, 150]$	625.6 ± 11.3	> 15.03	15.09 ± 0.11
	1611.2005	-2.8665	ESI	$[-100, 60]$	35.9 ± 9.6	15.09 ± 0.11	
	2344.2140	-0.9431	ESI	$[-150, 150]$	1043.0 ± 20.5	> 14.67	
	2374.4612	-1.5045	ESI	$[-150, 150]$	756.2 ± 23.5	> 15.07	
	2382.7650	-0.4949	ESI	$[-150, 150]$	1234.8 ± 22.6	> 14.30	
	2586.6500	-1.1605	ESI	$[-150, 150]$	1116.5 ± 34.9	> 14.84	
	2600.1729	-0.6216	ESI	$[-150, 150]$	1334.0 ± 21.8	> 14.38	
Ni II							
	1709.6042	-1.4895	ESI	$[-100, 100]$	86.9 ± 18.1	14.06 ± 0.09	14.02 ± 0.05
	1741.5531	-1.3696	ESI	$[-100, 100]$	97.5 ± 18.9	13.99 ± 0.08	
	1751.9157	-1.5575	ESI	$[-100, 100]$	66.6 ± 18.7	14.00 ± 0.11	
Zn II							
	2026.1360	-0.3107	ESI	$[-100, 100]$	69.7 ± 15.6	12.73 ± 0.09	12.73 ± 0.09
K4:J2343-1047 at $z = 2.68800$							
C I							
	1560.3092	-0.8808	ESI	$[-80, 80]$	< 19.8	< 13.03	< 13.03
	1656.9283	-0.8273	ESI	$[-80, 80]$	< 35.1	< 13.16	
C II							

Continued on Next Page...

Table 5.3 – Continued

Ion	λ (Å)	$\log f$	Instr.	$v_{\rm int}$ (km s^{-1})	W_λ (mÅ)	$\log N$	$\log N_{\rm adopt}$
	1334.5323	−0.8935	ESI	$[-80, 50]$	527.4 ± 17.2	> 14.82	> 14.82
C IV							
	1548.1950	−0.7194	ESI	$[-150, 150]$	490.6 ± 10.1	< 14.19	< 14.19
O I							
	1302.1685	−1.3110	ESI	$[-80, 60]$	504.8 ± 11.6	> 15.24	> 15.24
Al II							
	1670.7874	0.2742	ESI	$[-120, 90]$	328.9 ± 16.6	> 13.03	> 13.03
Si II							
	1260.4221	0.0030	ESI	$[-130, 80]$	534.2 ± 31.6	> 13.90	> 14.64
	1304.3702	−1.0269	ESI	$[-50, 80]$	329.9 ± 13.3	> 14.64	
	1526.7066	−0.8962	ESI	$[-100, 130]$	425.5 ± 5.5	> 14.47	
	1808.0130	−2.6603	ESI	$[-100, 100]$	< 45.5	< 15.04	
Cr II							
	2066.1610	−1.2882	ESI	$[-60, 60]$	< 33.9	< 13.45	< 13.44
Fe II							
	1608.4511	−1.2366	ESI	$[-100, 100]$	198.4 ± 23.4	14.29 ± 0.05	14.27 ± 0.03
	1611.2005	−2.8665	ESI	$[-80, 80]$	< 40.5	< 15.30	
	2260.7805	−2.6126	ESI	$[-100, 100]$	< 63.9	< 14.94	
	2344.2140	−0.9431	ESI	$[-130, 130]$	480.6 ± 27.3	> 14.13	
	2374.4612	−1.5045	ESI	$[-100, 80]$	238.1 ± 15.7	14.27 ± 0.03	
	2382.7650	−0.4949	ESI	$[-150, 150]$	711.2 ± 18.4	> 13.99	
	2600.1729	−0.6216	ESI	$[-150, 150]$	814.2 ± 77.7	> 14.07	
Ni II							
	1741.5531	−1.3696	ESI	$[-100, 100]$	< 39.4	< 13.71	< 13.71
	1751.9157	−1.5575	ESI	$[-100, 100]$	< 36.3	< 13.88	
Zn II							
	2026.1360	−0.3107	ESI	$[-100, 100]$	< 47.5	< 12.60	< 12.60

K5:J0343-0622 at $z = 2.57130$

Continued on Next Page...

Table 5.3 – Continued

Ion	λ (Å)	$\log f$	Instr.	$v_{\rm int}$ (km s^{-1})	W_λ (mÅ)	$\log N$	$\log N_{\rm adopt}$
C I							
	1656.9283	-0.8273	ESI	$[-60, 60]$	< 62.4	< 13.40	< 13.40
C II							
	1334.5323	-0.8935	ESI	$[-40, 30]$	206.9 ± 11.5	> 14.22	> 14.22
Al II							
	1670.7874	0.2742	ESI	$[-80, 60]$	152.2 ± 28.4	12.62 ± 0.08	12.62 ± 0.08
Al III							
	1854.7164	-0.2684	ESI	$[-80, 80]$	< 65.2	< 12.77	< 12.77
Si II							
	1260.4221	0.0030	ESI	$[-120, 80]$	701.2 ± 35.9	> 14.05	> 14.59
	1526.7066	-0.8962	ESI	$[-70, 100]$	564.0 ± 23.5	> 14.59	
	1808.0130	-2.6603	ESI	$[-80, 80]$	< 43.9	< 15.02	
Cr II							
	2056.2539	-0.9788	ESI	$[-80, 80]$	< 137.9	< 13.74	< 13.56
	2062.2340	-1.1079	ESI	$[-80, 80]$	< 69.2	< 13.56	
	2066.1610	-1.2882	ESI	$[-80, 80]$	< 63.8	< 13.70	
Mn II							
	2594.4990	-0.5670	ESI	$[-80, 80]$	< 112.3	< 13.04	< 13.04
Fe II							
	1608.4511	-1.2366	ESI	$[-50, 70]$	106.8 ± 12.7	13.98 ± 0.05	13.82 ± 0.04
	1611.2005	-2.8665	ESI	$[-80, 80]$	< 34.6	< 15.22	
	2260.7805	-2.6126	ESI	$[-80, 80]$	< 61.3	< 14.93	
	2344.2140	-0.9431	ESI	$[-60, 60]$	242.9 ± 28.6	13.75 ± 0.06	
	2374.4612	-1.5045	ESI	$[-40, 40]$	74.9 ± 22.3	13.75 ± 0.13	
	2382.7650	-0.4949	ESI	$[-50, 50]$	260.5 ± 22.4	> 13.35	
	2586.6500	-1.1605	ESI	$[-50, 40]$	203.2 ± 37.8	13.81 ± 0.09	
Ni II							
	1741.5531	-1.3696	ESI	$[-80, 80]$	< 49.9	< 13.83	< 13.83

Continued on Next Page...

Table 5.3 – Continued

Ion	λ (Å)	$\log f$	Instr.	$v_{\rm int}$ (km s^{-1})	W_λ (mÅ)	$\log N$	$\log N_{\rm adopt}$
Zn II							
	2026.1360	−0.3107	ESI	$[-80, 80]$	< 52.8	< 12.66	< 12.66
K6:J2351+1600 at $z = 3.78614$							
C I							
	1656.9283	−0.8273	ESI	$[-100, 100]$	< 47.0	< 13.32	< 13.32
C IV							
	1548.1950	−0.7194	ESI	$[-100, 130]$	105.9 ± 20.4	13.47 ± 0.08	13.47 ± 0.08
O I							
	1302.1685	−1.3110	ESI	$[-100, 60]$	363.7 ± 14.2	> 14.96	> 14.96
Al II							
	1670.7874	0.2742	ESI	$[-40, 40]$	126.6 ± 12.0	12.56 ± 0.05	12.56 ± 0.05
Al III							
	1862.7895	−0.5719	ESI	$[-80, 80]$	< 40.6	< 12.88	< 12.88
Si II							
	1190.4158	−0.6017	ESI	$[-60, 40]$	255.7 ± 12.3	> 14.23	> 14.23
	1260.4221	0.0030	ESI	$[-70, 50]$	358.7 ± 13.0	> 13.78	
	1526.7066	−0.8962	ESI	$[-60, 70]$	190.5 ± 22.6	> 14.10	
	1808.0130	−2.6603	ESI	$[-100, 70]$	< 50.7	< 15.10	
Fe II							
	1608.4511	−1.2366	ESI	$[-60, 60]$	100.4 ± 13.8	13.97 ± 0.06	13.97 ± 0.06
	1611.2005	−2.8665	ESI	$[-50, 50]$	57.5 ± 12.7	< 15.29	
Ni II							
	1454.8420	−1.4908	ESI	$[-80, 80]$	< 16.4	< 13.61	< 13.61
K7:J0042-1020 at $z = 2.75440$							

Continued on Next Page...

Table 5.3 – Continued

Ion	λ (Å)	$\log f$	Instr.	$v_{\rm int}$ (km s^{-1})	W_λ (mÅ)	$\log N$	$\log N_{\rm adopt}$
C I							
	1656.9283	−0.8273	ESI	$[-80, 80]$	< 20.2	< 12.93	< 12.93
C II							
	1334.5323	−0.8935	ESI	$[-100, 220]$	981.0 ± 11.0	> 15.07	> 15.07
C IV							
	1548.1950	−0.7194	ESI	$[-180, 180]$	915.7 ± 10.4	14.53 ± 0.01	14.53 ± 0.01
O I							
	1302.1685	−1.3110	ESI	$[-80, 160]$	729.9 ± 12.5	> 15.32	> 15.31
Al II							
	1670.7874	0.2742	ESI	$[-180, 200]$	732.2 ± 14.4	> 13.42	> 13.42
Al III							
	1854.7164	−0.2684	ESI	$[-100, 100]$	< 23.9	< 12.35	< 12.35
	1862.7895	−0.5719	ESI	$[-100, 100]$	< 24.8	< 12.65	
Si II							
	1260.4221	0.0030	ESI	$[-90, 200]$	962.4 ± 12.1	> 14.22	15.52 ± 0.03
	1526.7066	−0.8962	ESI	$[-80, 180]$	600.6 ± 10.1	> 14.56	
	1808.0130	−2.6603	ESI	$[-120, 180]$	186.8 ± 11.7	15.52 ± 0.03	
Si IV							
	1393.7550	−0.2774	ESI	$[-180, 180]$	636.8 ± 13.1	13.97 ± 0.01	13.97 ± 0.01
Cr II							
	2056.2539	−0.9788	ESI	$[-100, 100]$	< 34.5	< 13.15	< 13.15
	2066.1610	−1.2882	ESI	$[-100, 100]$	< 24.7	< 13.29	
Mn II							
	2594.4990	−0.5670	ESI	$[-80, 80]$	< 110.9	< 13.28	< 12.99
	2606.4620	−0.7151	ESI	$[-80, 80]$	< 74.7	< 12.99	
Fe II							
	1608.4511	−1.2366	ESI	$[-100, 190]$	309.9 ± 9.2	14.44 ± 0.01	14.39 ± 0.01
	1611.2005	−2.8665	ESI	$[-100, 190]$	< 20.3	< 14.99	

Continued on Next Page...

Table 5.3 – Continued

Ion	λ (Å)	$\log f$	Instr.	$v_{\rm int}$ (km s^{-1})	W_λ (mÅ)	$\log N$	$\log N_{\rm adopt}$
	2249.8768	−2.7397	ESI	$[-100, 190]$	< 39.0	< 14.85	
	2260.7805	−2.6126	ESI	$[-100, 190]$	< 37.1	< 14.70	
	2344.2140	−0.9431	ESI	$[-100, 190]$	627.1 ± 16.2	> 14.17	
	2374.4612	−1.5045	ESI	$[-100, 190]$	236.4 ± 19.7	14.23 ± 0.04	
	2382.7650	−0.4949	ESI	$[-100, 190]$	1062.3 ± 24.6	> 14.09	
	2600.1729	−0.6216	ESI	$[-100, 190]$	1042.9 ± 40.8	> 14.05	
Ni II							
	1709.6042	−1.4895	ESI	$[-100, 100]$	< 27.4	< 13.69	< 13.56
	1741.5531	−1.3696	ESI	$[-100, 100]$	< 27.6	< 13.57	
	1751.9157	−1.5575	ESI	$[-100, 100]$	< 34.4	< 13.83	
				K9:J0949+1115 at $z = 2.75840$			
C I							
	1656.9283	−0.8273	ESI	$[-100, 100]$	< 15.8	< 12.82	< 12.82
C II							
	1334.5323	−0.8935	ESI	$[-70, 100]$	762.3 ± 5.2	> 15.04	> 15.03
C IV							
	1548.1950	−0.7194	ESI	$[-80, 80]$	582.2 ± 4.4	> 14.45	> 14.52
	1550.7700	−1.0213	ESI	$[-80, 80]$	445.5 ± 4.9	> 14.52	
O I							
	1302.1685	−1.3110	ESI	$[-100, 100]$	885.1 ± 7.4	> 15.54	> 15.53
Al II							
	1670.7874	0.2742	ESI	$[-130, 130]$	777.7 ± 7.4	> 13.60	> 13.60
Al III							
	1854.7164	−0.2684	ESI	$[-100, 100]$	178.4 ± 8.7	13.09 ± 0.02	13.05 ± 0.02
	1862.7895	−0.5719	ESI	$[-100, 100]$	61.4 ± 9.0	12.90 ± 0.06	
Si II							

Continued on Next Page...

Table 5.3 – Continued

Ion	λ (Å)	$\log f$	Instr.	$v_{\rm int}$ (km s^{-1})	W_λ (mÅ)	$\log N$	$\log N_{\rm adopt}$
	1260.4221	0.0030	ESI	$[-150, 200]$	1106.5 ± 13.1	> 14.28	15.41 ± 0.02
	1304.3702	−1.0269	ESI	$[-100, 100]$	533.5 ± 9.0	> 14.95	
	1526.7066	−0.8962	ESI	$[-100, 100]$	772.3 ± 5.8	> 14.86	
	1808.0130	−2.6603	ESI	$[-100, 100]$	145.4 ± 7.7	15.41 ± 0.02	
Si IV							
	1393.7550	−0.2774	ESI	$[-120, 80]$	504.1 ± 7.0	> 13.99	> 14.08
	1402.7700	−0.5817	ESI	$[-80, 80]$	357.5 ± 6.9	> 14.08	
Cr II							
	2056.2539	−0.9788	ESI	$[-100, 100]$	80.2 ± 9.7	< 13.34	13.19 ± 0.13
	2062.2340	−1.1079	ESI	$[-100, 70]$	57.8 ± 10.1	< 13.32	
	2066.1610	−1.2882	ESI	$[-100, 100]$	28.4 ± 9.0	13.19 ± 0.13	
Fe II							
	1608.4511	−1.2366	ESI	$[-120, 120]$	522.3 ± 6.4	> 14.88	15.10 ± 0.05
	2260.7805	−2.6126	ESI	$[-100, 100]$	128.4 ± 13.2	15.10 ± 0.05	
	2344.2140	−0.9431	ESI	$[-100, 100]$	981.7 ± 8.2	> 14.62	
	2374.4612	−1.5045	ESI	$[-100, 100]$	896.2 ± 21.0	> 15.05	
	2382.7650	−0.4949	ESI	$[-100, 100]$	1176.0 ± 7.7	> 14.26	
	2586.6500	−1.1605	ESI	$[-100, 100]$	687.2 ± 40.0	> 14.53	
	2600.1729	−0.6216	ESI	$[-180, 180]$	1350.4 ± 27.2	> 14.34	
Ni II							
	1709.6042	−1.4895	ESI	$[-100, 100]$	50.5 ± 9.3	13.80 ± 0.08	13.72 ± 0.06
	1741.5531	−1.3696	ESI	$[-100, 100]$	45.0 ± 10.4	13.63 ± 0.10	
	1751.9157	−1.5575	ESI	$[-100, 100]$	40.3 ± 11.3	13.75 ± 0.12	
				K10:J1018+3106 at $z = 2.45917$			
C II							
	1334.5323	−0.8935	ESI	$[-170, 60]$	455.3 ± 10.8	> 14.70	> 14.70

Continued on Next Page...

Table 5.3 – Continued

Ion	λ (Å)	$\log f$	Instr.	$v_{\rm int}$ (km s^{-1})	W_λ (mÅ)	$\log N$	$\log N_{\rm adopt}$
C IV							
	1548.1950	−0.7194	ESI	$[-40, 110]$	421.2 ± 7.3	> 14.28	> 14.28
O I							
	1302.1685	−1.3110	ESI	$[-120, 80]$	480.0 ± 8.8	> 15.18	> 15.18
Al II							
	1670.7874	0.2742	ESI	$[-80, 80]$	296.4 ± 5.7	> 13.10	> 13.10
Al III							
	1854.7164	−0.2684	ESI	$[-70, 70]$	102.3 ± 8.0	12.86 ± 0.03	12.86 ± 0.03
	1862.7895	−0.5719	ESI	$[-70, 70]$	53.1 ± 7.9	12.85 ± 0.06	
Si II							
	1304.3702	−1.0269	ESI	$[-60, 60]$	288.0 ± 6.8	> 14.59	15.29 ± 0.03
	1526.7066	−0.8962	ESI	$[-70, 70]$	304.6 ± 7.0	> 14.38	
	1808.0130	−2.6603	ESI	$[-70, 70]$	108.4 ± 7.3	15.29 ± 0.03	
Si IV							
	1393.7550	−0.2774	ESI	$[-70, 70]$	302.6 ± 6.2	> 13.83	> 13.83
S II							
	1250.5840	−2.2634	ESI	$[-50, 50]$	< 22.7	< 14.66	< 14.65
Cr II							
	2062.2340	−1.1079	ESI	$[-50, 50]$	< 12.0	< 12.80	< 12.80
	2066.1610	−1.2882	ESI	$[-50, 50]$	< 12.1	< 12.96	
Mn II							
	2594.4990	−0.5670	ESI	$[-50, 50]$	33.8 ± 9.9	12.34 ± 0.13	12.34 ± 0.13
	2606.4620	−0.7151	ESI	$[-50, 50]$	< 21.0	< 12.44	
Fe II							
	1608.4511	−1.2366	ESI	$[-50, 40]$	180.3 ± 3.7	> 14.28	14.30 ± 0.03
	2249.8768	−2.7397	ESI	$[-50, 50]$	< 16.2	< 14.48	
	2260.7805	−2.6126	ESI	$[-50, 50]$	< 15.4	< 14.33	
	2344.2140	−0.9431	ESI	$[-50, 50]$	361.4 ± 7.8	> 14.09	
	2374.4612	−1.5045	ESI	$[-70, 50]$	240.6 ± 13.2	14.30 ± 0.03	
	2382.7650	−0.4949	ESI	$[-80, 60]$	500.6 ± 12.0	> 13.80	

Continued on Next Page...

Table 5.3 – Continued

Ion	λ (Å)	$\log f$	Instr.	$v_{\rm int}$ (km s^{-1})	W_λ (mÅ)	$\log N$	$\log N_{\rm adopt}$
	2586.6500	−1.1605	ESI	$[-50, 50]$	367.1 ± 10.3	> 14.19	
	2600.1729	−0.6216	ESI	$[-50, 50]$	510.8 ± 7.8	> 13.89	
Ni II							
	1709.6042	−1.4895	ESI	$[-50, 50]$	< 13.8	< 13.40	< 13.26
	1741.5531	−1.3696	ESI	$[-50, 50]$	< 13.8	< 13.26	
	1751.9157	−1.5575	ESI	$[-50, 50]$	< 13.3	< 13.42	
Zn II							
	2026.1360	−0.3107	ESI	$[-50, 50]$	< 13.7	< 12.07	< 12.07
				K11:J0851+2332 at $z = 3.52970$			
C I							
	1560.3092	−0.8808	ESI	$[-100, 100]$	< 24.9	< 13.12	< 13.04
	1656.9283	−0.8273	ESI	$[-100, 100]$	< 25.6	< 13.04	
C II							
	1334.5323	−0.8935	ESI	$[-200, 210]$	1826.6 ± 12.5	> 15.42	> 15.42
C IV							
	1548.1950	−0.7194	ESI	$[-270, 250]$	1144.3 ± 16.2	> 14.72	> 14.76
	1550.7700	−1.0213	ESI	$[-250, 250]$	811.7 ± 16.5	> 14.76	
O I							
	1302.1685	−1.3110	ESI	$[-190, 260]$	1768.4 ± 19.0	> 15.82	> 15.82
Al II							
	1670.7874	0.2742	ESI	$[-200, 250]$	1864.2 ± 13.5	> 14.00	> 14.00
Si II							
	1260.4221	0.0030	ESI	$[-190, 220]$	1670.7 ± 10.9	> 14.52	> 15.57
	1304.3702	−1.0269	ESI	$[-230, 260]$	1958.1 ± 17.1	> 15.57	
	1526.7066	−0.8962	ESI	$[-220, 250]$	1765.4 ± 11.0	> 15.22	
Fe II							

Continued on Next Page...

Table 5.3 – Continued

Ion	λ (Å)	$\log f$	Instr.	$v_{\rm int}$ (km s^{-1})	W_λ (mÅ)	$\log N$	$\log N_{\rm adopt}$
	1608.4511	−1.2366	ESI	$[-200, 250]$	1106.9 ± 19.0	> 15.15	> 15.15
	1611.2005	−2.8665	ESI	$[-180, 180]$	< 48.8	< 15.37	
Ni II							
	1709.6042	−1.4895	ESI	$[-200, 200]$	66.5 ± 16.1	13.92 ± 0.10	13.95 ± 0.07
	1741.5531	−1.3696	ESI	$[-200, 200]$	98.7 ± 22.2	13.97 ± 0.09	
	1751.9157	−1.5575	ESI	$[-200, 200]$	< 43.3	< 13.95	
Zn II							
	2026.1360	−0.3107	ESI	$[-160, 160]$	86.2 ± 25.9	12.73 ± 0.12	12.73 ± 0.12
				K12:J0956+1448 at $z = 2.66065$			
C I							
	1560.3092	−0.8808	ESI	$[-100, 100]$	< 15.7	< 12.92	< 12.85
	1656.9283	−0.8273	ESI	$[-100, 100]$	< 16.9	< 12.85	
C II							
	1334.5323	−0.8935	ESI	$[-90, 40]$	442.7 ± 5.9	> 14.71	> 14.71
C IV							
	1548.1950	−0.7194	ESI	$[-150, 60]$	181.5 ± 6.8	13.74 ± 0.02	13.72 ± 0.01
	1550.7700	−1.0213	ESI	$[-150, 50]$	78.6 ± 7.0	13.64 ± 0.04	
N I							
	1200.2233	−1.0645	ESI	$[-50, 50]$	191.1 ± 6.4	< 14.41	14.36 ± 0.03
	1200.7098	−1.3665	ESI	$[-50, 50]$	101.1 ± 7.1	14.36 ± 0.03	
O I							
	1302.1685	−1.3110	ESI	$[-120, 50]$	471.4 ± 6.0	> 15.14	> 15.14
Al II							
	1670.7874	0.2742	ESI	$[-120, 50]$	325.1 ± 6.4	> 13.09	> 13.09
Al III							
	1854.7164	−0.2684	ESI	$[-50, 60]$	46.1 ± 5.5	12.48 ± 0.05	12.46 ± 0.05

Continued on Next Page...

Table 5.3 – Continued

Ion	λ (Å)	$\log f$	Instr.	$v_{\rm int}$ (km s^{-1})	W_λ (mÅ)	$\log N$	$\log N_{\rm adopt}$
	1862.7895	−0.5719	ESI	$[-50, 60]$	20.8 ± 5.4	12.41 ± 0.11	
Si II							
	1526.7066	−0.8962	ESI	$[-100, 100]$	324.0 ± 5.5	> 14.40	14.90 ± 0.07
	1808.0130	−2.6603	ESI	$[-60, 50]$	46.8 ± 7.4	14.90 ± 0.07	
Si IV							
	1393.7550	−0.2774	ESI	$[-150, 50]$	319.6 ± 7.4	13.67 ± 0.01	13.67 ± 0.01
Cr II							
	2056.2539	−0.9788	ESI	$[-70, 80]$	28.9 ± 9.0	12.89 ± 0.13	12.95 ± 0.10
	2062.2340	−1.1079	ESI	$[-80, 80]$	< 16.5	< 12.94	
	2066.1610	−1.2882	ESI	$[-80, 80]$	< 16.4	13.11 ± 0.14	
Fe II							
	1608.4511	−1.2366	ESI	$[-70, 60]$	209.9 ± 6.5	> 14.38	14.61 ± 0.09
	2249.8768	−2.7397	ESI	$[-60, 20]$	< 19.8	< 14.57	
	2260.7805	−2.6126	ESI	$[-50, 20]$	42.4 ± 8.4	14.61 ± 0.09	
	2344.2140	−0.9431	ESI	$[-80, 50]$	418.4 ± 6.5	> 14.13	
	2382.7650	−0.4949	ESI	$[-100, 70]$	743.8 ± 7.3	> 14.03	
	2586.6500	−1.1605	ESI	$[-80, 50]$	422.0 ± 22.2	> 14.24	
	2600.1729	−0.6216	ESI	$[-80, 50]$	584.4 ± 19.9	> 13.93	
Ni II							
	1751.9157	−1.5575	ESI	$[-40, 70]$	31.1 ± 7.1	13.64 ± 0.10	13.64 ± 0.10
Zn II							
	2026.1360	−0.3107	ESI	$[-80, 80]$	35.2 ± 9.1	12.31 ± 0.11	12.31 ± 0.11
				K13:J1151+3536 at $z = 2.59780$			
C I							
	1656.9283	−0.8273	ESI	$[-100, 100]$	< 19.5	< 12.92	< 12.92
C II							

Continued on Next Page...

Table 5.3 – Continued

Ion	λ (Å)	$\log f$	Instr.	$v_{\rm int}$ (km s^{-1})	W_λ (mÅ)	$\log N$	$\log N_{\rm adopt}$
	1334.5323	−0.8935	ESI	$[-160, 130]$	946.1 ± 9.4	> 15.05	> 15.05
O I							
	1302.1685	−1.3110	ESI	$[-100, 80]$	809.0 ± 6.6	> 15.51	> 15.51
Al II							
	1670.7874	0.2742	ESI	$[-110, 110]$	619.9 ± 8.4	> 13.38	> 13.38
Al III							
	1854.7164	−0.2684	ESI	$[-100, 100]$	< 22.1	< 12.31	< 12.31
	1862.7895	−0.5719	ESI	$[-100, 100]$	< 20.9	< 12.59	
Si II							
	1304.3702	−1.0269	ESI	$[-60, 100]$	516.8 ± 7.6	> 14.90	15.13 ± 0.05
	1526.7066	−0.8962	ESI	$[-60, 120]$	711.1 ± 4.4	> 14.85	
	1808.0130	−2.6603	ESI	$[-70, 80]$	80.3 ± 9.3	15.13 ± 0.05	
Cr II							
	2066.1610	−1.2882	ESI	$[-80, 80]$	< 20.4	< 13.20	< 13.20
Fe II							
	1608.4511	−1.2366	ESI	$[-180, 180]$	401.1 ± 10.9	14.62 ± 0.01	14.61 ± 0.01
	1611.2005	−2.8665	ESI	$[-100, 100]$	< 18.1	< 14.95	
	2260.7805	−2.6126	ESI	$[-180, 200]$	< 34.8	< 14.68	
	2344.2140	−0.9431	ESI	$[-180, 200]$	954.3 ± 18.3	> 14.54	
	2374.4612	−1.5045	ESI	$[-110, 120]$	492.3 ± 11.3	14.61 ± 0.01	
	2382.7650	−0.4949	ESI	$[-180, 180]$	1252.3 ± 11.3	> 14.27	
	2586.6500	−1.1605	ESI	$[-120, 110]$	835.8 ± 23.9	> 14.53	
	2600.1729	−0.6216	ESI	$[-120, 130]$	1297.0 ± 20.2	> 14.35	
Zn II							
	2026.1360	−0.3107	ESI	$[-120, 100]$	61.8 ± 11.7	12.55 ± 0.08	12.55 ± 0.08
H1:J2123-0053 at $z = 2.78030$							

Continued on Next Page...

Table 5.3 – Continued

Ion	λ (Å)	$\log f$	Instr.	$v_{\rm int}$ (km s^{-1})	W_λ (mÅ)	$\log N$	$\log N_{\rm adopt}$
C I							
	1560.3092	-0.8808	ESI	$[-100, 100]$	< 33.4	< 13.26	< 13.00
	1656.9283	-0.8273	ESI	$[-100, 100]$	< 23.6	< 13.00	
C II							
	1334.5323	-0.8935	ESI	$[-110, 170]$	971.8 ± 14.3	> 15.07	> 15.06
O I							
	1302.1685	-1.3110	ESI	$[-50, 100]$	595.2 ± 10.3	> 15.34	> 15.34
Al II							
	1670.7874	0.2742	ESI	$[-150, 150]$	543.3 ± 13.1	> 13.29	> 13.29
Al III							
	1854.7164	-0.2684	ESI	$[-50, 50]$	66.5 ± 7.7	12.64 ± 0.05	12.65 ± 0.05
	1862.7895	-0.5719	ESI	$[-50, 50]$	35.4 ± 7.5	12.66 ± 0.09	
Si II							
	1260.4221	0.0030	ESI	$[-180, 180]$	852.3 ± 15.5	> 14.14	> 14.54
	1526.7066	-0.8962	ESI	$[-100, 160]$	509.4 ± 11.2	> 14.54	
	1808.0130	-2.6603	ESI	$[-60, 60]$	< 20.9	< 14.71	
Cr II							
	2056.2539	-0.9788	ESI	$[-50, 50]$	< 18.9	< 12.87	< 12.87
	2062.2340	-1.1079	ESI	$[-50, 50]$	< 28.7	< 13.16	
	2066.1610	-1.2882	ESI	$[-50, 50]$	< 22.8	< 13.27	
Fe II							
	1608.4511	-1.2366	ESI	$[-100, 120]$	200.5 ± 11.9	14.28 ± 0.03	14.27 ± 0.02
	2260.7805	-2.6126	ESI	$[-80, 120]$	< 29.0	< 14.60	
	2344.2140	-0.9431	ESI	$[-90, 140]$	492.3 ± 13.5	> 14.13	
	2374.4612	-1.5045	ESI	$[-80, 100]$	223.7 ± 14.7	14.26 ± 0.03	
	2382.7650	-0.4949	ESI	$[-100, 120]$	763.2 ± 15.0	> 13.97	
	2586.6500	-1.1605	ESI	$[-100, 100]$	319.0 ± 39.2	> 14.08	
	2600.1729	-0.6216	ESI	$[-100, 120]$	808.0 ± 43.3	> 13.97	
Ni II							
	1709.6042	-1.4895	ESI	$[-100, 100]$	< 26.1	< 13.66	< 13.61

Continued on Next Page...

Table 5.3 – Continued

Ion	λ (Å)	$\log f$	Instr.	$v_{\rm int}$ (km s^{-1})	W_λ (mÅ)	$\log N$	$\log N_{\rm adopt}$
	1741.5531	−1.3696	ESI	$[-100, 100]$	< 31.6	< 13.61	
	1751.9157	−1.5575	ESI	$[-100, 100]$	< 28.7	< 13.76	
H4:J0816+4823 at $z = 2.70665$							
C I							
	1656.9283	−0.8273	ESI	$[-100, 100]$	< 20.3	< 12.93	< 12.93
C II							
	1334.5323	−0.8935	ESI	$[-150, 150]$	606.8 ± 14.7	> 14.68	> 14.68
O I							
	1302.1685	−1.3110	ESI	$[-60, 60]$	236.0 ± 8.5	> 14.70	> 14.70
Al II							
	1670.7874	0.2742	ESI	$[-60, 60]$	103.6 ± 7.7	12.44 ± 0.03	12.44 ± 0.03
Si II							
	1193.2897	−0.3018	ESI	$[-30, 60]$	143.8 ± 11.1	> 13.57	> 13.85
	1260.4221	0.0030	ESI	$[-30, 60]$	279.0 ± 7.7	> 13.63	
	1526.7066	−0.8962	ESI	$[-60, 60]$	133.7 ± 5.0	> 13.85	
	1808.0130	−2.6603	ESI	$[-60, 50]$	< 17.4	< 14.62	
Mn II							
	2594.4990	−0.5670	ESI	$[-60, 60]$	< 53.3	< 12.68	< 12.68
	2606.4620	−0.7151	ESI	$[-60, 60]$	< 54.0	< 12.82	
Fe II							
	1611.2005	−2.8665	ESI	$[-50, 50]$	< 16.3	< 14.88	13.65 ± 0.02
	2249.8768	−2.7397	ESI	$[-50, 50]$	< 23.1	< 14.64	
	2260.7805	−2.6126	ESI	$[-50, 50]$	< 21.9	< 14.48	
	2344.2140	−0.9431	ESI	$[-60, 50]$	178.2 ± 7.9	13.63 ± 0.02	
	2374.4612	−1.5045	ESI	$[-40, 40]$	84.5 ± 7.3	13.78 ± 0.04	
	2382.7650	−0.4949	ESI	$[-40, 60]$	221.3 ± 8.7	> 13.31	

Continued on Next Page...

Table 5.3 – Continued

Ion	λ (Å)	$\log f$	Instr.	$v_{\rm int}$ (km s^{-1})	W_λ (mÅ)	$\log N$	$\log N_{\rm adopt}$
	2586.6500	-1.1605	ESI	$[-60, 60]$	155.3 ± 25.0	13.65 ± 0.07	
	2600.1729	-0.6216	ESI	$[-50, 50]$	233.2 ± 22.0	> 13.33	
Ni II							
	1709.6042	-1.4895	ESI	$[-60, 60]$	< 16.9	< 13.49	< 13.49
	1751.9157	-1.5575	ESI	$[-60, 60]$	< 20.4	< 13.62	
Zn II							
	2026.1360	-0.3107	ESI	$[-60, 60]$	< 17.3	< 12.16	< 12.16
				H10:J0751+4516 at $z = 2.68266$			
C I							
	1560.3092	-0.8808	ESI	$[-100, 100]$	< 32.4	< 13.23	< 12.99
	1656.9283	-0.8273	ESI	$[-100, 100]$	< 23.3	< 12.99	
C IV							
	1548.1950	-0.7194	ESI	$[-50, 70]$	99.7 ± 10.5	13.45 ± 0.05	13.44 ± 0.04
	1550.7700	-1.0213	ESI	$[-50, 70]$	44.0 ± 11.0	13.37 ± 0.11	
O I							
	1302.1685	-1.3110	ESI	$[-40, 30]$	236.2 ± 8.2	> 14.81	> 14.81
Al II							
	1670.7874	0.2742	ESI	$[-70, 60]$	169.9 ± 10.3	> 12.72	> 12.72
Al III							
	1854.7164	-0.2684	ESI	$[-50, 40]$	36.9 ± 10.2	12.38 ± 0.12	12.38 ± 0.12
	1862.7895	-0.5719	ESI	$[-50, 40]$	< 18.4	< 12.53	
Si II							
	1304.3702	-1.0269	ESI	$[-60, 50]$	246.8 ± 11.5	> 14.42	14.85 ± 0.09
	1526.7066	-0.8962	ESI	$[-60, 50]$	152.9 ± 9.6	> 13.98	
	1808.0130	-2.6603	ESI	$[-60, 50]$	42.2 ± 8.6	14.85 ± 0.09	
Si IV							

Continued on Next Page...

Table 5.3 – Continued

Ion	λ (Å)	$\log f$	Instr.	$v_{\rm int}$ (km s^{-1})	W_λ (mÅ)	$\log N$	$\log N_{\rm adopt}$
	1393.7550	−0.2774	ESI	$[-40, 40]$	123.0 ± 7.8	13.23 ± 0.03	13.23 ± 0.03
Cr II							
	2056.2539	−0.9788	ESI	$[-50, 50]$	< 24.0	< 12.97	> 13.04
	2062.2340	−1.1079	ESI	$[-50, 50]$	< 21.5	> 13.04	
Fe II							
	1608.4511	−1.2366	ESI	$[-50, 50]$	135.2 ± 12.1	14.14 ± 0.04	14.15 ± 0.03
	1611.2005	−2.8665	ESI	$[-50, 50]$	< 25.2	< 15.06	
	2249.8768	−2.7397	ESI	$[-50, 50]$	< 39.5	< 14.91	
	2260.7805	−2.6126	ESI	$[-50, 50]$	< 27.2	< 14.58	
	2344.2140	−0.9431	ESI	$[-60, 50]$	261.8 ± 17.3	> 13.88	
	2374.4612	−1.5045	ESI	$[-60, 50]$	174.0 ± 14.0	14.15 ± 0.04	
	2382.7650	−0.4949	ESI	$[-60, 50]$	285.9 ± 13.6	> 13.47	
	2600.1729	−0.6216	ESI	$[-60, 50]$	244.8 ± 40.6	> 13.57	
Ni II							
	1454.8420	−1.4908	ESI	$[-50, 50]$	< 12.6	< 13.49	< 13.49
	1709.6042	−1.4895	ESI	$[-50, 50]$	< 22.5	< 13.63	
	1741.5531	−1.3696	ESI	$[-50, 50]$	< 25.9	< 13.55	
	1751.9157	−1.5575	ESI	$[-50, 50]$	< 26.3	< 13.73	
Zn II							
	2026.1360	−0.3107	ESI	$[-50, 50]$	< 24.8	< 12.33	< 12.32
				H11:J0818+0720 at $z = 3.23320$			
C I							
	1560.3092	−0.8808	ESI	$[-200, 300]$	< 48.8	< 13.41	< 13.26
	1656.9283	−0.8273	ESI	$[-200, 300]$	< 43.8	< 13.26	
C II							
	1334.5323	−0.8935	ESI	$[-250, 300]$	2192.2 ± 14.6	> 15.47	> 15.47

Continued on Next Page...

Table 5.3 – Continued

Ion	λ (Å)	$\log f$	Instr.	$v_{\rm int}$ (km s^{-1})	W_λ (mÅ)	$\log N$	$\log N_{\rm adopt}$
C IV							
	1548.1950	−0.7194	ESI	$[-160, 320]$	749.9 ± 19.1	> 14.57	> 14.82
	1550.7700	−1.0213	ESI	$[-160, 320]$	657.9 ± 21.6	> 14.82	
O I							
	1302.1685	−1.3110	ESI	$[-250, 320]$	2206.8 ± 14.4	> 15.91	> 15.91
Al II							
	1670.7874	0.2742	ESI	$[-200, 320]$	1178.8 ± 19.7	> 13.69	> 13.69
Si II							
	1260.4221	0.0030	ESI	$[-100, 320]$	1423.2 ± 12.4	> 14.41	> 15.40
	1304.3702	−1.0269	ESI	$[-180, 300]$	1505.2 ± 14.3	> 15.41	
	1526.7066	−0.8962	ESI	$[-200, 320]$	1214.6 ± 14.4	> 14.99	
Si IV							
	1393.7550	−0.2774	ESI	$[-100, 230]$	1101.3 ± 18.7	> 14.46	> 14.46
Cr II							
	2056.2539	−0.9788	ESI	$[-200, 300]$	< 54.7	< 13.33	< 13.33
	2062.2340	−1.1079	ESI	$[-200, 300]$	157.5 ± 25.6	< 13.78	
	2066.1610	−1.2882	ESI	$[-200, 300]$	< 52.4	< 13.62	
Fe II							
	1608.4511	−1.2366	ESI	$[-200, 320]$	598.1 ± 18.7	> 14.82	> 15.20
	1611.2005	−2.8665	ESI	$[-200, 300]$	< 39.2	< 15.28	
	2249.8768	−2.7397	ESI	$[-200, 300]$	< 315.7	< 15.61	
	2344.2140	−0.9431	ESI	$[-140, 300]$	1593.9 ± 72.8	> 14.75	
	2374.4612	−1.5045	ESI	$[-150, 150]$	1023.0 ± 116.1	> 15.20	
Ni II							
	1709.6042	−1.4895	ESI	$[-150, 280]$	96.6 ± 23.3	14.09 ± 0.10	14.09 ± 0.07
	1741.5531	−1.3696	ESI	$[-150, 220]$	123.6 ± 30.2	14.08 ± 0.10	
Zn II							
	2026.1360	−0.3107	ESI	$[-180, 300]$	< 48.4	< 12.62	< 12.62

Continued on Next Page...

Table 5.3 – Continued

Ion	λ (Å)	$\log f$	Instr.	$v_{\rm int}$ (km s^{-1})	W_λ (mÅ)	$\log N$	$\log N_{\rm adopt}$
H13:J0811+3936 at $z = 2.65000$							
C I							
	1560.3092	−0.8808	ESI	$[-100, 200]$	< 27.0	< 13.15	< 13.15
	1656.9283	−0.8273	ESI	$[-100, 200]$	< 36.4	< 13.18	
C II							
	1334.5323	−0.8935	ESI	$[-100, 300]$	1074.3 ± 25.2	> 15.11	> 15.11
C IV							
	1548.1950	−0.7194	ESI	$[-100, 250]$	553.1 ± 13.7	> 14.43	> 14.51
	1550.7700	−1.0213	ESI	$[-100, 250]$	428.6 ± 13.7	> 14.51	
O I							
	1302.1685	−1.3110	ESI	$[-100, 300]$	1118.4 ± 15.8	> 15.53	> 15.53
Al II							
	1670.7874	0.2742	ESI	$[-100, 300]$	807.6 ± 17.2	> 13.41	> 13.41
Al III							
	1854.7164	−0.2684	ESI	$[-100, 300]$	132.4 ± 19.7	12.95 ± 0.06	12.98 ± 0.05
	1862.7895	−0.5719	ESI	$[-100, 300]$	98.0 ± 19.7	13.10 ± 0.09	
Si II							
	1190.4158	−0.6017	ESI	$[-200, 200]$	728.5 ± 29.2	> 14.68	> 14.96
	1193.2897	−0.3018	ESI	$[-200, 200]$	1391.3 ± 100.7	> 14.74	
	1260.4221	0.0030	ESI	$[-100, 300]$	1392.0 ± 24.3	> 14.39	
	1304.3702	−1.0269	ESI	$[-80, 260]$	756.9 ± 15.4	> 14.96	
	1526.7066	−0.8962	ESI	$[-100, 300]$	803.6 ± 17.5	> 14.68	
	1808.0130	−2.6603	ESI	$[-40, 230]$	< 39.8	< 14.99	
Si IV							
	1393.7550	−0.2774	ESI	$[-120, 200]$	434.4 ± 12.6	> 13.87	> 13.95
	1402.7700	−0.5817	ESI	$[-120, 200]$	295.7 ± 14.3	> 13.95	
Cr II							
	2056.2539	−0.9788	ESI	$[-100, 200]$	< 39.7	< 13.19	< 13.19

Continued on Next Page...

Table 5.3 – Continued

Ion	λ (Å)	$\log f$	Instr.	$v_{\rm int}$ (km s^{-1})	W_λ (mÅ)	$\log N$	$\log N_{\rm adopt}$
	2062.2340	-1.1079	ESI	$[-100, 200]$	< 54.5	< 13.45	
	2066.1610	-1.2882	ESI	$[-100, 200]$	< 39.1	< 13.48	
Fe II							
	1608.4511	-1.2366	ESI	$[-120, 280]$	304.1 ± 23.6	14.42 ± 0.03	14.45 ± 0.03
	1611.2005	-2.8665	ESI	$[-100, 200]$	< 42.0	< 15.31	
	2249.8768	-2.7397	ESI	$[-100, 250]$	< 83.6	< 15.16	
	2260.7805	-2.6126	ESI	$[-100, 250]$	< 73.5	< 15.01	
	2344.2140	-0.9431	ESI	$[-100, 300]$	865.2 ± 27.6	> 14.33	
	2374.4612	-1.5045	ESI	$[-60, 250]$	414.7 ± 39.6	14.50 ± 0.04	
	2382.7650	-0.4949	ESI	$[-100, 300]$	1268.8 ± 31.7	> 14.14	
Ni II							
	1709.6042	-1.4895	ESI	$[-100, 200]$	< 28.4	< 13.71	< 13.62
	1741.5531	-1.3696	ESI	$[-100, 200]$	< 31.8	< 13.62	
	1751.9157	-1.5575	ESI	$[-100, 200]$	< 35.1	< 13.85	
Zn II							
	2026.1360	-0.3107	ESI	$[-100, 200]$	< 52.7	< 12.66	< 12.66

Chapter 6

Testing models for molecular gas formation in galaxies: hydrostatic pressure or gas and dust shielding?

This chapter reproduces the paper "Testing models for molecular gas formation in galaxies: hydrostatic pressure or gas and dust shielding?" by Fumagalli et al. (2010a). This work offers the theoretical framework to interpret the results presented in Fumagalli et al. (2012). In modern theories of star formation, the key factor that triggers the formation of stars is the ability of the gas to shield from the ambient UV radiation field and cool. Because the thermodynamic conditions that are favorable for star formation are also ideal for the conversion of atomic hydrogen into molecular hydrogen (Krumholz et al. 2011, Glover & Clark 2012), H_2 becomes an effective tracer for star formation, as evident from the empirical correlation between molecular gas surface density and star formation rate surface density in local galaxies (e.g. Bigiel et al. 2008). According to these theories, the lack of significant *in situ* star formation in damped Lyman$-\alpha$ systems is a consequence of the lower metal content and lower hydrogen column

density in these systems (Krumholz et al. 2009c). Before applying these models that have been tested in local galaxies to the high redshift Universe, we wish to verify whether they can be applied to environments similar to those found in high redshift galaxies. This test is the subject of this chapter.

6.1 Abstract

Stars in galaxies form in giant molecular clouds that coalesce when the atomic hydrogen is converted into molecules. There are currently two dominant models for what property of the galactic disk determines its molecular fraction: either hydrostatic pressure driven by the gravity of gas and stars, or a combination of gas column density and metallicity. To assess the validity of these models, we compare theoretical predictions to the observed atomic gas content of low-metallicity dwarf galaxies with high stellar densities. The extreme conditions found in these systems are optimal to distinguish the two models, otherwise degenerate in nearby spirals. Locally, on scales < 100 pc, we find that the state of the interstellar medium is mostly sensitive to the gas column density and metallicity rather than hydrostatic pressure. On larger scales where the average stellar density is considerably lower, both pressure and shielding models reproduce the observations, even at low metallicity. We conclude that models based on gas and dust shielding more closely describe the process of molecular formation, especially at the high resolution that can be achieved in modern galaxy simulations or with future radio/millimeter arrays.

6.2 Introduction

Theoretical arguments based on gravitational instability as well as observations of molecular gas reveal that low temperature ($T \sim 10$ K) and high density ($n \sim 40$ cm^{-3}) giant

molecular clouds (GMCs) are the natural sites where stars form. Although individual GMCs can be resolved only in the Milky Way or in a handful of local galaxies (e.g. Bolatto et al. 2008; and references therein), CO observations of several nearby spirals show that star formation mostly occurs in molecular regions[1] (e.g. Wong & Blitz 2002, Kennicutt et al. 2007, Bigiel et al. 2008). At the same time, neutral atomic hydrogen (H I) remains the primordial constituent of the molecular phase (H_2), playing an essential role in the formation of new stars, as shown by the low star-formation rate (SFR) (e.g. Boselli & Gavazzi 2006) and low molecular content (Fumagalli et al. 2009) found in H I-poor galaxies. Therefore, the transition from H I to H_2 is a key process that drives and regulates star formation in galaxies.

The problem of molecule formation has been studied extensively in the literature mainly through two different approaches. The first is by modelling the formation of molecular gas empirically, starting from CO and H I maps in nearby galaxies. Following this path, Wong & Blitz (2002; hereafter WB02), Blitz & Rosolowsky (2004; hereafter BR04), and Blitz & Rosolowsky (2006; hereafter BR06) have inferred that the molecular (Σ_{H_2}) to atomic (Σ_{HI}) surface density ratio

$$R_{H_2} = \Sigma_{H_2}/\Sigma_{HI} \tag{6.1}$$

in disks is a function solely of the hydrostatic midplane pressure P_m, which is driven both by the stellar and gas density: $R_{H_2} \sim P_m^{0.92}$ (hereafter BR model). The second approach models the microphysics that regulates the formation of H_2 and its photodissociation. A detailed description should take into account the balance of H_2 formation onto dust grains and its dissociation by Lyman-Werner (LW) photons and cosmic rays, together with a complete network of chemical reactions that involves several molecules generally found in the interstellar medium (ISM). Due to this complexity, many studies address mainly the detailed physics of H_2 in individual clouds,

[1] Whether stars form from regions entirely dominated by molecules at the outskirts of disks remains an open question due to difficulties in probing molecular hydrogen via common tracers such as CO in those environments (e.g. Fumagalli & Gavazzi 2008, Leroy et al. 2008).

without considering molecular formation on galactic scales (e.g. the pioneering work by van Dishoeck & Black 1986).

Elmegreen (1993) made an early attempt to produce a physically-motivated prescription for molecule formation in galaxies by studying the H I to H_2 transition in both self-gravitating and diffuse clouds as a function of the external ISM pressure P_e and radiation field intensity j. This numerical calculation shows that the molecular fraction $f_{H_2} = \Sigma_{H_2}/\Sigma_{gas} \sim P_e^{2.2} j^{-1}$, with $\Sigma_{gas} = \Sigma_{H_2} + \Sigma_{HI}$. More recently, properties of the molecular ISM have been investigated with hydrodynamical simulations by Robertson & Kravtsov (2008) who have concluded that the H_2 destruction by the interstellar radiation field drives the abundance of molecular hydrogen and empirical relations such as the R_{H_2}/P_m correlation. Using numerical simulations which include self-consistent metal enrichment, Gnedin et al. (2009) and Gnedin & Kravtsov (2010) have stressed also the importance of metallicity in regulating the molecular fraction and therefore the SFR. Similarly, Pelupessy et al. (2006) developed a subgrid model to track in hydrodynamical simulations the formation of H_2 on dust grains and its destruction by UV irradiation in the cold gas phase and collisions in the warm gas phase.

A different approach based entirely on first principles has been proposed in a series of papers by Krumholz et al. (2008; hereafter KMT08), Krumholz et al. (2009a; hereafter KMT09), and McKee & Krumholz (2010; hereafter MK10). Their model (hereafter KMT model) describes the atomic-to-molecular transition in galaxies using a physically motivated prescription for the dust shielding and self-shielding of molecular hydrogen. This work differs from previous analyses mainly because it provides an analytic expression for f_{H_2} as a function of the total gas column density and metallicity (Z). Therefore, the KMT model can be used to approximate the molecular gas on galactic scales without a full radiative transfer calculation.

In this paper we shall consider the BR and the KMT models as examples of the two different approaches used to describe the H I to H_2 transition in galaxies. Remarkably,

both formalisms predict values for $f_{\rm H_2}$ which are roughly consistent with atomic and molecular observations in local disk galaxies (see Krumholz et al. 2009a; sect. 4.1.3). The reason is that the BR model becomes dependent on the gas column density alone if the stellar density is fixed to typical values found in nearby galaxies (see the discussion in Sect. 6.6). Despite the observed agreement, there are significant conceptual differences: the BR model is empirical and does not address the details of the ISM physics, while the KMT model approximates physically motivated prescriptions for the H_2 formation as functions of observables. Hence, although in agreement for solar metallicity at resolutions above a few hundred parsecs, the two prescriptions may not be fully equivalent in different regimes. It is still an open question whether molecule formation is mainly driven by hydrostatic pressure or UV radiation shielding over different spatial scales and over a large range of metallicities.

A solution to this problem has important implications in several contexts. From a theoretical point of view, cosmological simulations of galaxy formation that span a large dynamic range will benefit from a simple prescription for molecular gas formation in order to avoid computationally intense radiative transfer calculations. Similarly, semi-analytic models or post-processing of dark-matter-only simulations will greatly benefit from a simple formalism that describes the molecular content in galaxies. Observationally, the problem of understanding the gas molecular ratio has several connections with future radio or millimeter facilities (e.g. ALMA, the Atacama Large Millimeter Array or SKA, the Square Kilometer Array). In fact, these interferometers will allow high-resolution mapping of atomic and molecular gas across a large interval of redshift and galactic locations over which metallicity and intensity of the local photodissociating UV radiation vary significantly.

In this work, we explore the validity of the KMT and BR models in nearby dwarf starbursts both locally (< 100 pc) and on larger scales (~ 1 kpc). Their low metallicity (down to a few hundredths of solar values) combined with the relatively high stellar densities found

in these systems offers an extreme environment in which the similarity between the two models breaks down. In fact, for a fixed gas column density, high stellar density corresponds to high pressure and therefore high molecular fraction in the BR model. Conversely, for a fixed gas column density, low metallicity in the KMT model results in a low molecular fraction (see Figure 6.1).

We emphasize that the BR model was not designed to describe the molecular fraction on scales smaller than several hundred parsecs; indeed, Blitz & Rosolowsky (2006) explicitly warn against applying their model on scales smaller than about twice the pressure scale height of a galaxy. However, a number of theoretical models have extrapolated the BR model into this regime (e.g. Narayanan et al. 2010, Murante et al. 2010). The analysis we present at small scales (< 100 pc, comparable to the resolutions of simulations in which the BR model has been used) is aimed at highlighting the issues that arise from such an extrapolation. Furthermore, a comparison of the pressure and shielding (KMT) models across a large range of physical scales offers additional insight into the physical processes responsible for the atomic-to-molecular transition.

The paper is organized as follows: after a brief review of the two models in Section 6.3, we will present two data-sets collected from the literature in Section 6.4. The comparison between models and observations is presented in Section 6.5, while discussion and conclusions follow in Sections 6.6 and 6.7. Throughout this paper we assume a solar photospheric abundance $12 + \log(\mathrm{O/H}) = 8.69$ from Asplund et al. (2009). Also, we make use of dimensionless gas surface densities $\Sigma_{\rm gas}^{'} = \Sigma_{\rm gas}/(\mathrm{M_\odot pc^{-2}})$, stellar densities $\rho_{\rm star}^{'} = \rho_{\rm star}/(\mathrm{M_\odot pc^{-3}})$, gas velocity dispersions $v_{\rm gas}^{'} = v_{\rm gas}/(\mathrm{km\ s^{-1}})$, and metallicity $Z' = Z/Z_\odot$.

6.3 Models

Here we summarize the basic concepts of the BR and KMT models which are relevant for our discussion. The reader should refer to the original works (WB02, BR04, BR06, KMT08, KMT09, and MK10) for a complete description of the two formalisms.

6.3.1 The BR model

The *ansatz* at the basis of the BR model is that the molecular ratio $R_{\rm H_2}$ is entirely determined by the midplane hydrostatic pressure $P_{\rm m}$ according to the power-law relation:

$$R_{\rm H_2} = \left(\frac{P_{\rm m}}{P_\circ}\right)^{\alpha} \ . \tag{6.2}$$

The pressure can be evaluated with an approximate solution for the hydrostatic equilibrium for a two-component (gas and stars) disk (Elmegreen 1989):

$$P_{\rm m} \sim G\Sigma_{\rm gas}\left(\Sigma_{\rm gas} + \Sigma_{\rm star}\frac{v_{\rm gas}}{v_{\rm star}}\right) \ , \tag{6.3}$$

where $\Sigma_{\rm star}$ is the stellar density, and $v_{\rm star}$ and $v_{\rm gas}$ are the stellar and gas velocity dispersion, respectively.

For a virialized disk, equation (6.3) reduces to

$$P_{\rm m} \sim \sqrt{G}\Sigma_{\rm gas}v_{\rm gas}\left(\Sigma_{\rm gas}^{0.5}h_{\rm gas}^{-0.5} + \Sigma_{\rm star}^{0.5}h_{\rm star}^{-0.5}\right) \ ; \tag{6.4}$$

by replacing the surface densities with volume densities using the disk scale-heights h, the

previous equation can be rewritten as

$$P_{\rm m} \sim \sqrt{G}\Sigma_{\rm gas} v_{\rm gas} \left(\rho_{\rm gas}^{0.5} + \rho_{\rm star}^{0.5}\right) \ . \tag{6.5}$$

Under the assumption that $\rho_{\rm star} > \rho_{\rm gas}$, the contribution of the gas self-gravity can be neglected and equation (6.5) reduces to

$$\frac{P_{\rm m}}{k} = 272 \ \Sigma_{\rm gas}^{'} \ (2\rho_{\rm star})^{'0.5} \ v_{\rm gas}^{'} \ {\rm K \ cm^{-3}} \ . \tag{6.6}$$

where $\Sigma_{\rm gas}^{'} = \Sigma_{\rm gas}/({\rm M_\odot pc^{-2}})$, $\rho_{\rm star}^{'} = \rho_{\rm star}/({\rm M_\odot pc^{-3}})$ and $v_{\rm gas}^{'} = v_{\rm gas}/({\rm km \ s^{-1}})$. In deriving equation (6.6), constants in cgs units have been added to match equation (5) in BR06. Combining equation (6.2) with equation (6.6), the molecular ratio in the BR model becomes

$$R_{\rm H_2} = \left[8.95 \times 10^{-3} \ \Sigma_{\rm gas}^{'} \ \rho_{\rm star}^{'0.5} \ v_{\rm gas}^{'}\right]^{0.92} \ , \tag{6.7}$$

where the best fit values $P_\circ = (4.3 \pm 0.6) \times 10^4$ K cm^{-3} and $\alpha = 0.92$ have been derived from CO and H I observations of local spiral galaxies (BR06).

6.3.2 The KMT model

The core of the KMT model is a set of two coupled integro-differential equations for the radiative transfer of LW radiation and the balance between the H_2 phtodissociation and its formation onto dust grains. Neglecting the H_2 destruction by cosmic rays, the combined transfer-dissociation equation is

$$\vec{\nabla} \cdot \vec{F}^* = -n\sigma_{\rm d} c E^* - \frac{f_{\rm HI} n^2 \mathcal{R}}{f_{\rm diss}} \ . \tag{6.8}$$

On the left-hand side, $\vec{F}^*$ is the photon number flux integrated over the LW band. The first term on the right-hand side accounts for dust absorption, where n is the number density of hydrogen atoms, $\sigma_{\rm d}$ the dust cross section per hydrogen nucleus to LW-band photons, and E^* the photon number density integrated over the LW band. The second term accounts for absorption due to photodissociation, expressed in term of the H_2 formation rate in steady state conditions. Here, $f_{\rm HI}$ is the hydrogen atomic fraction while $f_{\rm diss}$ is the fraction of absorbed radiation that produces dissociation rather than a de-excitation into a newly bound state. Finally, $\mathcal{R}$ expresses the formation rate of molecular hydrogen on dust grains.

Equation (6.8) can be integrated for a layer of dust and a core of molecular gas mixed with dust. This solution specifies the transition between a fully atomic layer and a fully molecular core and hence describes the molecular fraction in the system as a function of the optical depth at which $F^* = 0$. Solutions to equation (6.8) can be rewritten as a function of two dimensionless constants

$$\tau_{\rm R} = n\sigma_{\rm d}R \tag{6.9}$$

and

$$\chi = \frac{f_{\rm diss}\sigma_{\rm d}cE^*}{n\mathcal{R}}\,. \tag{6.10}$$

Here, $\tau_{\rm R}$ is the dust optical depth for a cloud of size R, hence equation (6.9) specifies the dimensions of the system. Conversely, χ is the ratio of the rate at which the LW radiation is absorbed by dust grains to the rate at which it is absorbed by molecular hydrogen.

To introduce these equations which govern the microphysics of the H_2 formation into a formalism that is applicable on galactic scales, one has to assume (Wolfire et al. 2003) that a cold-neutral medium (CNM) is in pressure equilibrium with a warm-neutral medium (WNM). Assuming further that dust and metals in the gas component are proportional to the total metallicity ($Z' = Z/Z_\odot$, in solar units), equations (6.9) and (6.10) can be rewritten as a function

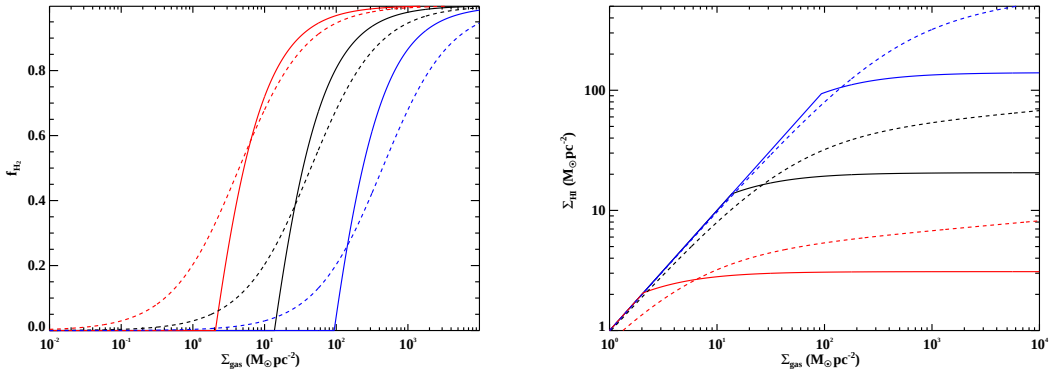

Figure 6.1: Left panel. Molecular fractions computed for the KMT (solid lines) and BR (dashed lines) models. Different lines represent three metallicities for the KMT model (from right to left, $Z' = 0.1$ blue, $Z' = 1$ black, and $Z' = 10$ red) or three stellar densities for the BR model (from right to left, $\rho'_{\rm star} = 0.001$ blue, $\rho'_{\rm star} = 0.1$ black, and $\rho'_{\rm star} = 10$ red). For the KMT model, we assume a clumping factor $c = 1$. Blue compact dwarfs (BCDs) at low metallicities and high stellar densities are the optimal systems to disentangle between the two models which are degenerate in massive spiral galaxies with solar metallicity (compare the two black lines). Right panel. Models for the H I surface density as a function of the total gas column density, for the same parameters adopted in the left panel. While the KMT model exhibits a well defined saturation in the atomic hydrogen, in the BR model $\Sigma_{\rm HI}$ increases asymptotically with $\Sigma_{\rm gas}$.

of the observed metallicity and gas surface density. Using the improved formalism described in MK10, the analytic approximation for the molecular fraction as specified by the solutions of equation (6.8) can be written as:

$$f_{\rm H_2} \simeq 1 - \left(\frac{3}{4}\right) \frac{s}{1 + 0.25s} \tag{6.11}$$

for $s < 2$ and $f_{\rm H_2} = 0$ for $s \geq 2$. Here $s = \ln(1+0.6\chi+0.01\chi^2)/(0.6\tau_c)$, $\chi = 0.76(1+3.1Z'^{0.365})$, and $\tau_c = 0.066\Sigma'_{\rm comp}Z'$. Finally, $\Sigma'_{\rm comp} = c\Sigma'_{\rm gas}$ where the clumping factor $c \geq 1$ is introduced to compensate for averaging observed gas surface densities over scales larger than the typical scale of the clumpy ISM. Primed surface densities are in units of $M_\odot$ pc^{-2}.

6.3.3 Differences between the two models

In Figure 6.1 we compare the BR and KMT models to highlight some behaviours that are relevant to our analysis. In the left panel, we present molecular fractions computed using the KMT (solid lines) and BR (dashed lines) formalisms. Different lines reflect three choices of metallicity for the KMT model (from right to left, $Z' = 0.1$ blue, $Z' = 1$ black, and $Z' = 10$ red) and three stellar densities for the BR model (from right to left, $\rho^{'}_{\rm star} = 0.001$ blue, $\rho^{'}_{\rm star} = 0.1$ black, and $\rho^{'}_{\rm star} = 10$ red). For a typical spiral disk with stellar mass $M_{\rm star} = 10^{10}{\rm M}_{\odot}$, size $R = 10$ kpc, and stellar height $h = 300$ pc, the stellar density is of the order of $\rho^{'}_{\rm star} \sim 0.1$. Figure 6.1 shows that, at solar metallicity and for a typical gas surface density $\Sigma^{'}_{\rm gas} \sim 10 - 100$, the two models predict similar molecular fractions.

To break the degeneracy, we apply model predictions to observations of blue compact dwarf galaxies (BCDs) or low-metallicity dwarf irregulars (dIrrs), characterized by high stellar density ($\rho^{'}_{\rm star} \sim 1 - 100$) and low metallicity ($Z' = 0.3 - 0.03$). In these environments, for a fixed gas surface density (excluding the limit $f_{\rm H_2} \rightarrow 1$), the two models predict very different molecular to atomic ratios.

In the right panel of Figure 6.1, we show the predicted atomic gas surface density as a function of the total gas surface density, for the same parameters selected in the left panel. Besides the dependence on the metallicity and stellar density, this plot reveals a peculiar difference between the two models. The KMT formalism exhibits a well defined saturation threshold in $\Sigma_{\rm HI}$ for a fixed value of metallicity. This corresponds to the maximum H I column density that is required to shield the molecular complex from the LW-band photons. All the atomic hydrogen that exceeds this saturation level is converted into molecular gas. Conversely, the BR model has no saturation in the atomic gas surface density, but it increases slowly as the total gas surface density increases.

6.4 The Dwarf Galaxy Samples

We study the behaviour of the KMT and the BR models using two data sets compiled from the literature. Specifically, we have selected low-metallicity compact dwarf galaxies with sufficient observations to constrain gas densities, stellar masses, and metal abundances for a comparison with models. The first sample comprises 16 BCDs and dIrrs, for which quantities integrated over the entire galaxy are available. These objects constitute a low-resolution sample with which we study the two models on galactic scales (> 1 kpc). For seven of these galaxies, we also have high-resolution H I maps and *Hubble Space Telescope* (*HST*) optical images. With these objects, we construct a high-resolution sample, useful to study the two formalisms at the scale of individual star cluster complexes (< 100 pc).

Both models depend on the total gas surface density. In principle, we could use CO emission, available in the form of integrated fluxes from the literature, to quantify $\Sigma_{\rm gas}$ and the molecular content of individual galaxies. However, recent studies of molecular hydrogen traced through a gas-to-dust ratio (e.g. Imara & Blitz 2007, Leroy et al. 2009a) support the idea that CO is a poor tracer of molecular hydrogen in low metallicity environments, mostly due to its inability to self-shield (Wolfire et al. 2010). Therefore, CO seems an unreliable H_2 tracer for these metal-poor galaxies. For this unfortunate reason, we avoid any attempt to precisely quantify $\Sigma_{\rm H_2}$, but rather use the observed H I column density as a lower limit on the total gas column densities.

As discussed in Section 6.3.3, the KMT model has a well-defined saturation threshold for $\Sigma_{\rm HI}$, and this threshold constitutes an observationally-testable prediction. The BR model does not have such a threshold and at a given $\rho_{\rm star}$ is in principle capable of producing arbitrarily high values of $\Sigma_{\rm HI}$ provided that the total gas density $\Sigma_{\rm gas}$ is sufficiently high. However, the extremely weak variation of $\Sigma_{\rm HI}$ with $\Sigma_{\rm gas}$ at large total gas column density ($\Sigma_{\rm HI} \propto \Sigma_{\rm gas}^{0.08}$)

means that the amount of total gas required to produce a given $\Sigma_{\rm HI}$ may be implausibly large. This effect allows us to check the BR model as well using only H I (Section 6.5), albeit not as rigorously as we can test the KMT model. We also check the robustness of our results in Appendix 6.8.3, where we impose an upper limit on $\Sigma_{\rm H_2}$ either from SFRs, assuming a depletion time $t_{\rm depl} \sim 2$ Gyr (Bigiel et al. 2008) for molecular gas, or from CO fluxes, using a conservative CO-to-H_2 conversion.

In the next sections, we discuss in detail the procedures adopted to derive gas surface densities and stellar densities for the two samples. The reader not interested in these rather technical aspects can find the analysis, discussion and conclusions starting from Section 6.5.

6.4.1 High-resolution sample

Seven BCDs are found in the literature with high-resolution H I maps and with sufficient ancillary *HST* data to infer stellar masses on scales < 100 pc, typical of individual GMCs. A detailed description of how we compute $\Sigma_{\rm HI}$ and $\rho_{\rm star}$ in individual galaxies is provided in Appendix 6.8.1, together with a list of relevant references. Here, we only summarize the general procedures we use.

Stellar masses of individual clusters are in a few cases directly taken from the literature. Otherwise, we infer stellar masses from integrated light by comparing two methods. The first is based on age estimates, whenever those are available in the literature. In this case, we infer stellar masses from observed absolute magnitudes by comparing the K or V band luminosity with predictions at the given age by Starburst99 (SB99; Leitherer et al. 1999). This is done assuming an instantaneous burst, similar metallicity, and a Salpeter initial mass function (IMF) with lower and upper mass limits at 1 and 100 $M_\odot$, respectively. The second method is based on optical and near-infrared colors (e.g., $B-V$, $V-I$, $V-J$, $V-H$, $V-K$). In this case, we use mass-to-light (M/L) ratios inferred from colors (Bell & de Jong 2001), and the stellar

masses are derived directly from observed luminosities. Usually, the two methods give similar results to within a factor of ~ 2.

Once the masses are known, we obtain stellar densities with sizes taken from the literature. If not available, we measure them by fitting a two-dimensional elliptical Gaussian to the clusters in *HST* images. For the closest objects, in order to avoid resolving individual stars, we fit binned surface brightness profiles with a one-dimensional Gaussian.

Stellar masses are probably the most uncertain quantities in our study. In fact, our first method suffers from the rapid changes in the broadband output of a starburst at young ages ($4-10$ Myr), due to the onset of red supergiants whose amplitude and time of onset depends on metallicity. Moreover, the initial mass function of the SB99 models and the lower-mass cutoff may introduce additional uncertainty, up to a factor of 3, considering a full range of systematic uncertainties (Bell et al. 2003). Instead, sources of error in the second method are the strong contribution of nebular continuum and line emission to the broadband colors of young starbursts. This can be a particularly severe problem in the K band because of recombination lines and free-free emission which in some cases constitutes as much as 50% of the broadband K magnitude (see Vanzi et al. 2000, Hunt et al. 2001; 2003). Despite this rather large uncertainty on the stellar densities, the results presented in the next sections can be considered rather robust. In fact, a variation in the density larger than the uncertainty would be required to significantly alter our conclusions. A more extensive discussion on this issue is presented in Section 6.5.1.2.

To complete the data set, we add to the gas and stellar densities values for the metallicity, distances, and SFR indicators as collected from the literature. Individual references are provided in Appendix 6.8.1 and Tables 6.1 and 6.2. We derive integrated SFRs using Hα, 60μm and radio free-free fluxes as different tracers. The final rates are given assuming the empirical calibrations by Kennicutt (1998) for the Hα, by Hopkins et al. (2002) for the 60μm, and by Hunt et al. (2005a) for the radio free-free emission. We note that this last tracer is optimum in

the absence of non-thermal emission, as typical in young starbursts. Since SFRs are used to set an upper limit on the total gas density assuming a given depletion time (Appendix 6.8.3), we choose the maximum value whenever more than one indicator is found for a single galaxy. Total SFR surface densities are then calculated adopting the galaxy sizes from NED[2]. A summary of the collected and derived data is presented in Table 6.1.

[2]NASA/IPAC Extragalactic Database.

Table 6.1: Data set for the high-resolution sample[a]

Name	Distance (Mpc)	12+log(O/H)	M^b_{star} ($10^5 M_\odot$)	Radius (pc)	ρ^b_{star} ($M_\odot pc^{-3}$)	Σ^b_{star} ($M_\odot pc^{-2}$)	Σ_{HI} ($10^{21} cm^{-2}$)	Σ_{sfr} ($M_\odot yr^{-1} kpc^{-2}$)
IZw18	13	7.19	4.7	56	0.64	47.9	3.5	0.134
			2.3	56	0.31	23.3		
SBS0335-052	53.7	7.23	10	18.2	39.5	959	7	0.154
			11		43.4	1055		
			4		15.8	384		
			11		43.4	1055		
			18		71.0	1726		
			2		7.9	192		
Mrk71	3.44	7.90	0.12	7.2	7.7	74	6	0.214
			0.054		3.5	33		
UM462	15.3	7.98	3.5-7.2	21	9-19	255-520	5.7	0.060
			1.2-2.2	27.5	1.4-2.5	51-92		
			2.1	43	0.6	36		
			1.8	27.5	2	74		
			1.6-2.9	33	1-2	48-85		
			1.6	50	0.3	21		
IIZw40	10.3	8.13	6.4-12	10.1	149-280	2005-3759	7.9	4.610
			1.3-15	5.2	222-2550	1536-17730		
NGC5253	3.5	8.19	0.13	3.0	115	460	6.4	0.181
			0.7-4	1.6-2.9	3915-4080	8704-15140		
			10-13	3.5	5570-7240	25980-33780		
NGC1140	18.2	8.20	9.1	7.3	558	5436	2.5	0.024
			59	6.6	4899	43114		

[a] Individual references are provided in Appendix 6.8.1, together with a detailed description of how quantities are measured. Metallicities and SFRs are the same as those given in Table 6.2.

[b] For the most uncertain values, we report the upper and lower limits. See Appendix 6.8.1 for further details.

The stellar densities in the high-resolution sample are generally quite high, and associated with massive compact star clusters, some of which are in the Super Star Cluster (SSC) category (e.g., O'Connell et al. 1994, Meurer et al. 1995, Whitmore et al. 2005). Despite their extreme properties, none of the BCDs in the high-resolution sample exceeds the maximum stellar surface density limit found by Hopkins et al. (2010), and most are $5-10$ times below this limit. Interestingly, the stellar densities here are uncorrelated with metallicity, implying that some other parameter must play the main role in defining the properties of massive star clusters.

6.4.2 Low-resolution sample

We have collected a second sample from the literature by requiring only that quantities integrated over the entire galaxy be available. Due to the lower spatial resolution, this data set is suitable to study the KMT and BR models on larger scales ($>$ 1 kpc). Our search yielded a total of 16 low-metallicity star-forming galaxies; among these are the 7 objects in the high-resolution sample. We have compiled gas and stellar densities, distances, and metallicity for these 16 objects, most of which are classified as BCDs, but some are dIrrs (Sm, Im), since they are more diffuse, larger in size, and more luminous (massive) than typical BCDs.

Stellar masses are computed from *Spitzer*/IRAC fluxes following the formulation of Lee et al. (2006), as we summarize in Appendix 6.8.2, together with a comment on the dominant sources of uncertainty. Stellar densities are then derived assuming spherical symmetry and the sizes inferred from the stellar component, as measured from IRAC images. The resolution of these images ($\sim 1\farcs2$) is a factor of 10 lower than the worst *HST* resolution, so that the compact regions are unresolved. This implies that the stellar densities derived for this sample are much lower than the values quoted for individual star cluster complexes. Moreover, for non-spherical (spheroidal) BCDs these densities correspond formally to lower limits; the volume of a prolate spheroid is smaller than the volume of a sphere by a factor $(b/a)^{0.5}$, with a and b the semi-

major and semi-minor axis, respectively. In our sample, the mean axis-ratio is $a/b \sim 1.5$ with a 0.5 standard deviation. This discrepancy is small enough to justify our assumption of spherical symmetry. In any case, the possible volume overestimate could partially compensate the potential overestimate of stellar density because of nebular emission contamination or free-free emission (see the discussion in Appendix 6.8.2).

For most objects, integrated H I fluxes are retrieved from HyperLeda[3] (Paturel et al. 2003a). We then convert integrated fluxes into mean column densities using optical radii from NED, and assuming that the gas extends twice as far as the stellar component (see Lee et al. 2002, van Zee et al. 1998a, Thuan et al. 2004). For I Zw 18, the integrated H I flux is not available in HyperLeda and we consider the flux published in de Vaucouleurs et al. (1991). Similarly, for Mrk 71 we estimate the total atomic gas from available interferometric observations averaged over the entire galaxy (Thuan et al. 2004).

^{12}CO$(1-0)$ fluxes are available for most of the galaxies here considered (see Table 6.2). For three galaxies, the most metal-poor objects in our sample (SBS 0335−052 E, I Zw 18, and Mrk 71), we find only CO upper limits in the literature. Because we use CO fluxes only to set upper limits on Σ_{H_2} (Appendix 6.8.3), we choose one of the largest CO-to-H_2 conversion measured to date (Leroy et al. 2009a). It is worth noting that for extremely metal poor galaxies (e.g. 1 Zw 18) the adopted conversion factor may still underestimate the H_2 content. To make our limits even more conservative, we compare these values with Σ_{H_2} inferred from SFRs and we choose for each galaxy the maximum of the two.

As with the high-resolution sample, we derive SFRs from Hα, 60μm, or free-free emission (see Table 6.2 for references). Again, SFR densities are computed assuming the optical size as given by NED. Finally, we collect information on the metallicity and distances for each object. A summary of the data derived for the low-resolution sample is given in Table 6.2.

[3] http://leda.univ-lyon1.fr

Table 6.2: Data set for the low-resolution sample

Name	Distance (Mpc)	12+log(O/H)[a]	Diameter[b] (kpc)	M_{star} (min-avg) $\log M_\odot$	Σ^c_{HI} ($M_\odot pc^{-2}$)	CO flux[d] (K km/s)	Σ^e_{sfr} ($M_\odot yr^{-1} kpc^{-2}$)	SFR tracer
Haro3	16.8	8.30	5.23 - 4.79	9.09 - 9.35	10.2	1.98	0.106	60μm
IIZw40	10.3	8.12	2.80 - 1.04	8.09 - 8.34	91.4	0.50	4.610	ff
IZw18	13.0	7.19	1.44 - 0.93	6.51 - 6.88	39.3	<1.00	0.134	ff
Mrk209	5.4	7.81	1.16 - 1.09	7.29 - 7.36	14.2	0.45	0.051	Hα
Mrk33	24.9	8.45	7.61 - 6.86	9.59 - 9.79	4.3	6.21	0.103	60μm
Mrk71	3.4	7.90	0.81 - 0.81	6.54 - 6.89	12.4	<0.34	0.214	Hα
NGC1140	18.2	8.20	6.56 - 6.55	9.30 - 9.58	15.7	0.97	0.024	Hα
NGC1156	7.1	8.23	5.53 - 5.93	8.62 - 9.10	5.0	0.76	0.007	Hα
NGC1741	55.1	8.05	17.82 -22.44	9.29 - 9.81	9.7	1.53	0.040	60μm
NGC2537	8.0	8.19	4.48 - 3.72	9.14 - 9.26	5.4	0.56	0.013	Hα
NGC4214	3.3	8.20	2.91 - 7.19	8.66 - 8.67	2.3	0.90	0.003	Hα
NGC5253	3.5	8.19	3.29 - 3.14	8.63 - 9.05	2.7	0.73	0.181	ff
NGC7077	13.3	8.04	2.72 - 2.90	8.54 - 8.60	3.6	0.68	0.014	Hα
SBS0335-052	53.7	7.23	4.24 - 3.35	7.79 - 8.72	24.5	<5.43	0.154	Hα
UM448	81.2	8.00	15.67 - 8.18	10.30 - 10.69	34.1	0.82	0.669	60μm
UM462	15.3	7.97	2.48 - 2.44	7.95 - 8.20	16.1	0.55	0.060	60μm

[a] References for metallicity: Davidge (1989); Campos-Aguilar et al. (1993); Izotov et al. (1997); Gil de Paz et al. (2002); Izotov & Thuan (2004a); Thuan & Izotov (2005); van Zee & Haynes (2006).

[b] Diameters as computed from stellar profiles and derived from NED.

[c] References for H I: HyperLeda except for I Zw 18 and Mrk 71, as described in the text.

[d] References for CO: Sage et al. (1992); Leon et al. (1998); Taylor et al. (1998); Barone et al. (2000); Dale et al. (2001); Gil de Paz et al. (2002); Albrecht et al. (2004); Leroy et al. (2005); Leroy et al. (2007).

[e] References for SFR: Young et al. (1989); Drissen et al. (2000); Hopkins et al. (2002); Gil de Paz et al. (2002); Hunt et al. (2004); Izotov & Thuan (2004a); Hunt et al. (2005a); Hunt et al. (2005b); Schmitt et al. (2006); van Zee & Haynes (2006).

6.5 Analysis

6.5.1 Testing models on small scales (< 100 pc)

With the aim of testing how the BR and KMT models perform at high stellar density and low metallicity, we first compare both formalisms with the observed H I surface densities and stellar densities in the high-resolution data set. Since the average resolution of this sample is below 100 pc, this part of the analysis focuses mainly on the molecular fraction in individual GMCs and associations rather than on larger ISM spatial scales. As previously mentioned, the BR formalism was not developed to describe the molecular fraction in such small regions (BR06). Hence the results presented in this section are intended to assess possible pitfalls of extrapolating the pressure model to small scales. Furthermore, a comparison of the performances of the BR and KMT models below 100 pc provides insight into what quantities are relevant to the production of molecules on different size scales.

As summarized in Section 6.3.2, the KMT formalism describes the molecular fraction as a function of the total gas column density and metallicity. A free parameter is the clumping factor c that maps the observed column density $\Sigma_{\rm gas}$ onto the relevant quantity in the model, i.e. the column density of the cold phase in individual clouds. With resolutions coarser than $\sim$100 pc, beam smearing dilutes the density peaks and one must adopt $c > 1$ in order to recover the intrinsic gas surface density $\Sigma_{\rm comp} = c\Sigma_{\rm gas}$. However, given the high resolution of the *HST* images, we set $c = 1$ so that the KMT model has no free parameters.

Conversely, as reviewed in Section 6.3.1, the BR model describes $R_{\rm H_2}$ as function of the total column density and stellar volume density. An additional parameter in this case is the gas velocity dispersion, set to $v'_{\rm gas} = 8$. Apart from a similar dependence on $\Sigma_{\rm gas}$, a direct comparison between models and observations is not straightforward. We start our analysis by confronting each model with observations, and then attempt a comparison of both models and

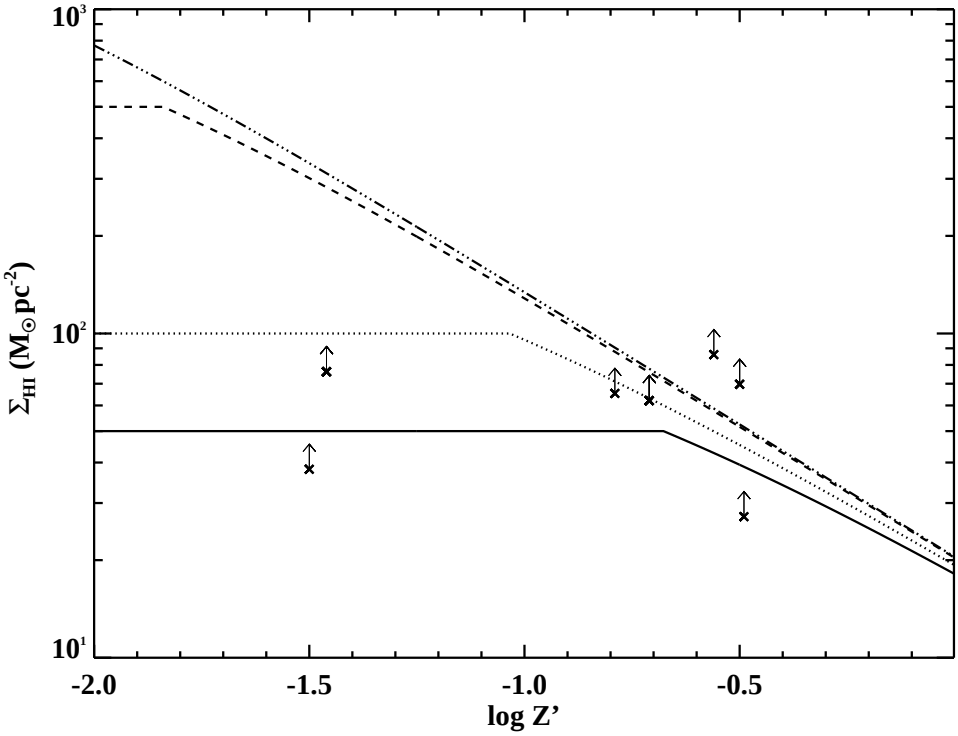

Figure 6.2: Comparison between observations of the H I surface density (crosses) with prediction from the KMT model (lines) as a function of the metallicity. Different curves are computed for different total gas column densities ($\Sigma'_{\rm gas} = 50$ solid line, $\Sigma'_{\rm gas} = 10^2$ dotted line, $\Sigma'_{\rm gas} = 5 \times 10^2$ dashed line, and $\Sigma'_{\rm gas} = 10^3$ dash-dotted line). The KMT model is computed for $c = 1$ and H I data are shown as lower limits since they are averaged over regions which are more extended than 100 pc (see the text for further details). Observations do not rule out the KMT model since there is a significant overlap between the parameter space allowed by both the theoretical curves and the data.

data.

6.5.1.1 The KMT model predictions below 100 pc

In Figure 6.2 we present the observed H I surface density (crosses) together with predictions from the KMT model (lines) as a function of the metallicity. Different curves are computed for different total gas column densities ($\Sigma'_{\rm gas} = 50$ solid line, $\Sigma'_{\rm gas} = 10^2$ dotted line, $\Sigma'_{\rm gas} = 5 \times 10^2$ dashed line, and $\Sigma'_{\rm gas} = 10^3$ dash-dotted line). Here, and for the rest of this

analysis, we correct the gas column density for helium with a standard coefficient 1.36. We do not include corrections for projection effects because in dwarf galaxies a unique inclination angle is not well defined for a warped (non-planar) H I distribution or in a triaxial system.

As discussed in Appendix 6.8.1, interferometric H I observations do not achieve the resolution required to match the *HST* observations. A possible solution is to downgrade *HST* images to match the atomic hydrogen maps. However, since the exact value for c would be unknown at the resultant resolution ($\gtrsim$ 100 pc), we perform our analysis on scales $<$ 100 pc, compatible with *HST* images and where $c \to 1$. For this reason, we express the observed H I column density as lower limits on the local $\Sigma_{\rm HI}$. This is because coarser spatial resolutions most likely average fluxes on larger areas, thus lowering the inferred peak column density. Indeed, whenever H I observations at different resolutions are compared, better resolution is associated with higher column densities. SBS 0335−052 E is an example: this BCD has $\Sigma_{\rm HI} = 7.4 \times 10^{20}$ cm^{-2} in a beam of $20\farcs5 \times 15\farcs0$ (Pustilnik et al. 2001) and 2×10^{21} cm^{-2} in a $3\farcs4$ beam (Ekta et al. 2009). The inferred H I column density is even higher, 7×10^{21} cm^{-2}, with the smaller $2''$ beam in *HST*/GHRS observations of Lyα absorption (Thuan & Izotov 1997).

In any case, the lower limits illustrated in Figure 6.2 prevent us from concluding that model and observations are in complete agreement, although this is strongly suggested. Adopting a conservative approach, this comparison shows that observations do not immediately rule out the KMT model on scales of $<$ 100 pc; 5/7 of the galaxies here considered are consistent with predicted curves. Although not crucial for the current and remaining analysis, the quoted metallicity may in some cases overestimate the dust and metal content which contributes to the H_2 formation. In fact, in the KMT model, it is the CNM that plays a relevant role in regulating $f_{\rm H_2}$ and the assumption that the nebular metallicity reflects the metal abundances in the cold ISM may not hold in all cases. Specifically, the optically-inferred metallicity used here is dominated by the ionized phase. Studies of the metal enrichment of the neutral gas

in metal poor dwarfs show that the neutral phase can be sometimes less metal-enriched than the ionized medium (e.g. Thuan et al. 2005, Lebouteiller et al. 2004, Lecavelier des Etangs et al. 2004). Furthermore, galaxies with the lowest nebular metallicities have similar neutral gas abundances, while dwarfs with higher ionized nebular metallicities can have up to $\sim$7 times (Lebouteiller et al. 2009) lower neutral ISM abundances (see however Bowen et al. 2005). A recent interpretation for this effect is that although mixing is effective in diffusing new metals from ionized regions, due to the larger volume, the enrichment is modest (Lebouteiller et al. 2009). This justifies the use of a single metallicity for multiple GMCs since, in the worst case, we would overestimate by some factor the local metal content. Data points, especially the ones at higher metallicity, would be offset to lower values and the parameter space common to data and model would increase, mitigating the discrepancy found at $Z' \sim -0.5$.

6.5.1.2 The BR model predictions below 100 pc

Turning our attention to the BR model, we test predictions (lines) against the observed $\Sigma_{\rm HI}$ (crosses) in Figure 6.3. Stellar densities are available for individual GMCs measured from high-resolution *HST* images. Therefore, while multiple observations overlap in metallicity in Figure 6.2, here we show distinct data-points for the same galaxy. Once again, different curves are for a selection of total gas density ($\Sigma'_{\rm gas} = 50$ solid line, $\Sigma'_{\rm gas} = 10^2$ dotted line, $\Sigma'_{\rm gas} = 5\times10^2$ dashed line, $\Sigma'_{\rm gas} = 10^3$ dash-dotted line, and $\Sigma'_{\rm gas} = 10^4$ long-dashed line); observed H I column densities are represented as lower limits on the local atomic gas column density. When stellar densities are particularly uncertain (see discussion in Appendix 6.8.1), we plot both lower and upper limits connected with a dotted line.

According to the BR model, despite the low metallicity, a high fraction of hydrogen is expected to be molecular because of the enhanced stellar density. However, Figure 6.3 illustrates that observations discourage the use of pressure models on scales $<$ 100 pc. Even under the

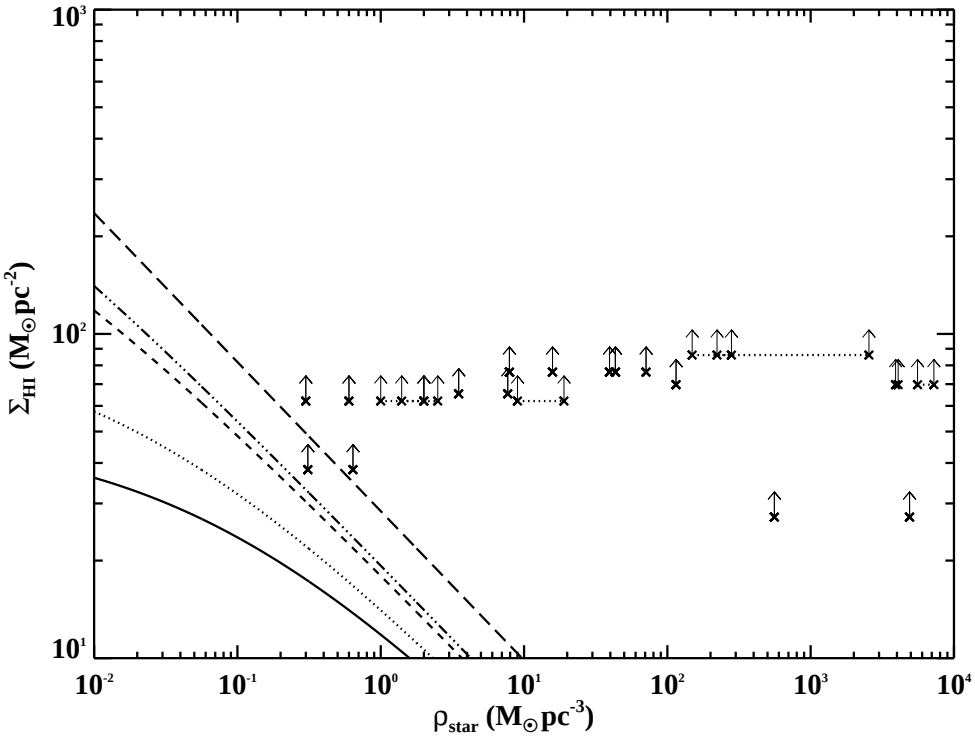

Figure 6.3: Comparison between the predicted $\Sigma_{\rm HI}$ from the BR model (lines) and observations (crosses) as a function of the stellar density in individual GMCs and associations. Different curves are for choices of total gas density ($\Sigma^{'}_{\rm gas} = 50$ solid line, $\Sigma^{'}_{\rm gas} = 10^2$ dotted line, $\Sigma^{'}_{\rm gas} = 5 \times 10^2$ dashed line, $\Sigma^{'}_{\rm gas} = 10^3$ dash-dotted line, and $\Sigma^{'}_{\rm gas} = 10^4$ long-dashed line). Observed H I column densities are represented as lower limits on the local atomic gas column density (see the text for further details). Observations rule out the BR model once extrapolated below 100 pc.

very conservative hypothesis that stellar densities are overestimated by a factor of $2-3$ and that the total column densities can reach very high values (e.g. the long-dashed curve at $\Sigma_{\rm gas}^{'} = 10^4$), observations mostly lie in the region not allowed by the extrapolation of the BR model[4]. Compatibility between the extrapolation of the BR model and a good fraction of the data would require $\Sigma_{\rm gas} \gtrsim 10^{10}$ M$_\odot$ pc^{-2}. A value that large would correspond to $A_V > 10^5$ with a dust-to-gas ratio that is 1% of the Milky Way value, and is thus ruled out by the fact that the star clusters are observable. Moreover, such a large $\Sigma_{\rm gas}$ would make the gas mass in the observed region larger than either the baryonic or the dark matter mass of the entire dwarf galaxy. Clearly, even though we cannot directly detect the molecular component of the gas, we can rule out the presence of such a large amount of gas on other grounds, and we can therefore conclude that the extrapolated BR model is incompatible with the observations. We give a more rigorous estimate of the maximum plausible value of $\Sigma_{\rm gas}^{\prime}$ in Appendix C.

An additional tunable parameter in the BR model is the gas velocity dispersion and a substantial change in $v_{\rm gas}$ can affect its predictions. In this paper, following BR04 and BR06, we adopt $v_{\rm gas}^{'} = 8$. Since pressure varies linearly with the velocity dispersion, we can solve for the value of $v_{\rm gas}$ required for the BR model to match the observations. We find that $v_{\rm gas}^{'} \lesssim 2$ in order to have one half of the data points consistent with the model; this is in contrast with recent H I observations (e.g. Chung et al. 2009, Walter et al. 2008) that show typical dispersion velocities $v_{\rm gas}^{'} > 5$ (and in many cases $v_{\rm gas}^{'} > 10$) in all the surveyed galaxies. It is worth mentioning that the use of the observed $v_{\rm gas}$ is not always appropriate, although unavoidable; for example, $v_{\rm gas}$ depends on the thermal velocity and the gas in a cold medium has a lower velocity dispersion than what inferred from a multiphase ISM. We conclude that the disagreement found in Figure 6.3 cannot be explained with uncertainties on the velocity dispersion.

[4]One might attempt to improve the agreement data-model by artificially smearing the stellar density down to the same resolution as the H I observations (while ignoring the presence of additional stars outside the *HST* PSF). This brings the total gas surface density required to match the majority of the observations down to $\Sigma_{\rm gas}^{\prime} \sim 10^3 - 10^4$, still a very large value.

Finally, we should assess if the high pressure predicted by the model can be attributed to the use of hydrostatic equilibrium in a disk rather than in a sphere, which would be more appropriate for our systems. Intuitively, this is not the case since the central pressure in a sphere of gas and stars cannot be lower than the midplane pressure of the disk. In fact, the central point in a sphere has to support the weight of the entire system, while each point in the midplane of a disk has to support only the pressure from the components along the vertical direction. This argument is substantiated by a quantitative analysis. A solution of the hydrostatic equilibrium equation for a sphere of gas and stars shows that the central gas pressure is enhanced by a quantity that depends on $(v_{\rm star}/v_{\rm gas})^2$. Therefore, to minimize an increase in the pressure due to the stellar component, the condition $(v_{\rm star}/v_{\rm gas}) \sim 1$ has to be satisfied. However, an increase in $v_{\rm gas}$ is reflected by an increase in the gas pressure itself. In other words, the central pressure in a gas sphere with cold kinematics (low $v_{\rm gas}$) receives a significant contribution from the stellar potential ($v_{\rm star}/v_{\rm gas} > 1$), while a gas sphere with hot kinematics (high $v_{\rm gas}$) has an intrinsically higher gas pressure ($v_{\rm star}/v_{\rm gas} < 1$).

6.5.1.3 A direct comparison between the two models

A different way to visualize both the models and the observations for the high-resolution sample is shown in Figure 6.4, where we plot predictions for $\Sigma_{\rm HI}$ of the KMT model (solid lines) and the BR model (dashed lines) as a function of the total gas column density. Different curves in the KMT model are for the maximum, central, and minimum metallicity observed in the sample (from the bottom to the top, $Z' = 0.32$ red, $Z' = 0.12$ black, and $Z' = 0.03$ blue). The curves for the BR model correspond to the maximum, central, and minimum stellar density (from the bottom to the top, $\log \rho^{'}_{\rm star} = 3.86$ red, $\log \rho^{'}_{\rm star} = 1.38$ black, and $\log \rho^{'}_{\rm star} = -0.52$ blue). As in the previous figures, observed lower limits on $\Sigma_{\rm HI}$ are superimposed. The yellow-shaded region in Figure 6.4 indicates the maximum parameter space allowed by the observations. This plot

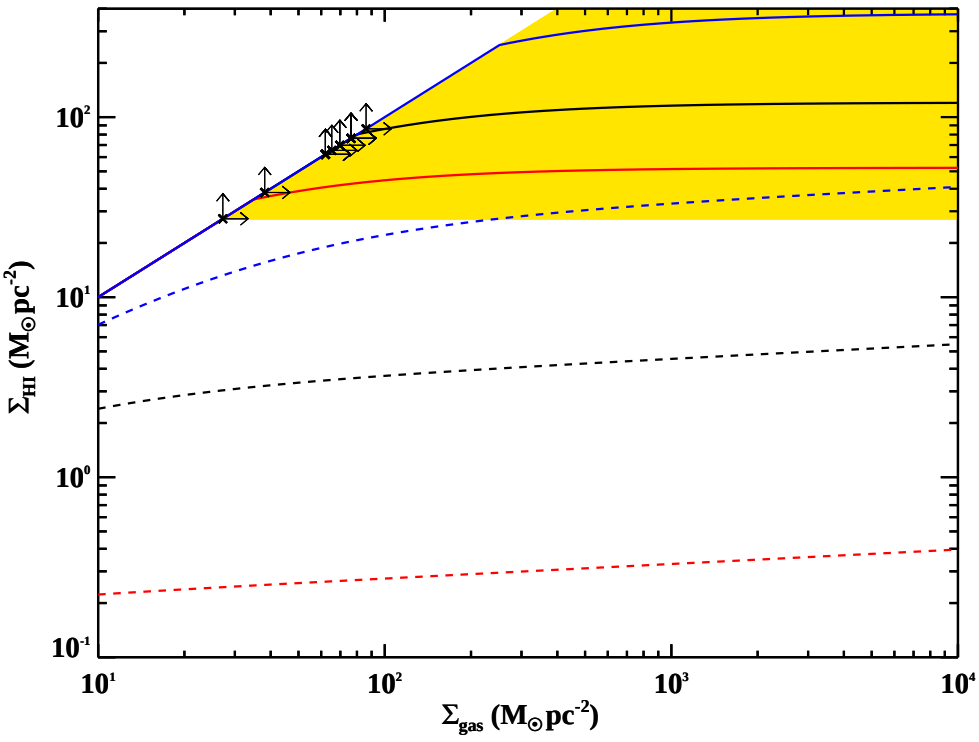

Figure 6.4: KMT model (solid lines) and BR model (dashed lines) as a function of the total gas density. Different curves are for the maximum, central, and minimum observed values of metallicity (in the KMT model: $Z' = 0.32$ red, $Z' = 0.12$ black, and $Z' = 0.03$ blue) and stellar density (in the BR model: $\log \rho'_{\rm star} = 3.86$ red, $\log \rho'_{\rm star} = 1.38$ black, and $\log \rho'_{\rm star} = -0.52$ blue). Observed lower limits on $\Sigma_{\rm HI}$ are superimposed. The yellow shaded region encloses the largest parameter-space allowed by observations. Data rule out any extrapolation of the BR model below 100 pc, where it largely overestimates the molecular fraction. Conversely, observations do not rule out immediately the KMT model.

summarizes the two main results presented in the previous paragraphs. Observations of $\Sigma_{\rm HI}$ reveal that an extrapolation of the BR model below scales of 100 pc results in a significant overestimation of the molecular fraction. In fact, for the observed $\Sigma_{\rm HI}$, exceedingly high total gas surface densities ($\Sigma^{'}_{\rm gas} > 10^4$) are required by the BR model to reproduce observations. As shown quantitatively in Appendix 6.8.3, such high values appear to be unrealistic. Conversely, observations seem to suggest a good agreement between the KMT model and data. Also, comparing Figure 6.4 with Figure 6.1, it appears that the different behaviour of the two models for a similar gas column density is related to which quantity regulates the molecular fraction at the second order, subordinately to the gas column density. In fact, the discrepancy with observations and the BR model is associated with the high values of stellar densities, while the consistency between the observed H I column densities and the KMT model is fostered by the low value of metallicity that raises the atomic hydrogen saturation limit.

6.5.2 Testing models on galactic scales (> 1 kpc)

In the second part of this analysis, we compare predictions from models and observations on larger scales (> 1 kpc), by considering spatially integrated quantities for a larger sample of BCDs. Before we start, it is worth mentioning that the condition $\Sigma^{'}_{\rm star} \gtrsim 20$ (see BR04) which ensures the validity of equation (6.6) holds also for the low-resolution data set.

6.5.2.1 A comparison between models and global data

In Figure 6.5 we present a comparison between observed H I surface densities and models, as previously done in Figure 6.4 for the high-resolution sample. Solid lines represent the KMT model, for the maximum ($Z' = 0.58$ red), central ($Z' = 0.20$ black), and minimum ($Z' = 0.03$ blue) observed metallicity. Dashed lines are for the BR model for the maximum ($\log \rho^{'}_{\rm star} = -1.45$ red), central ($\log \rho^{'}_{\rm star} = -2.00$ black), and minimum ($\log \rho^{'}_{\rm star} = -3.18$ blue)

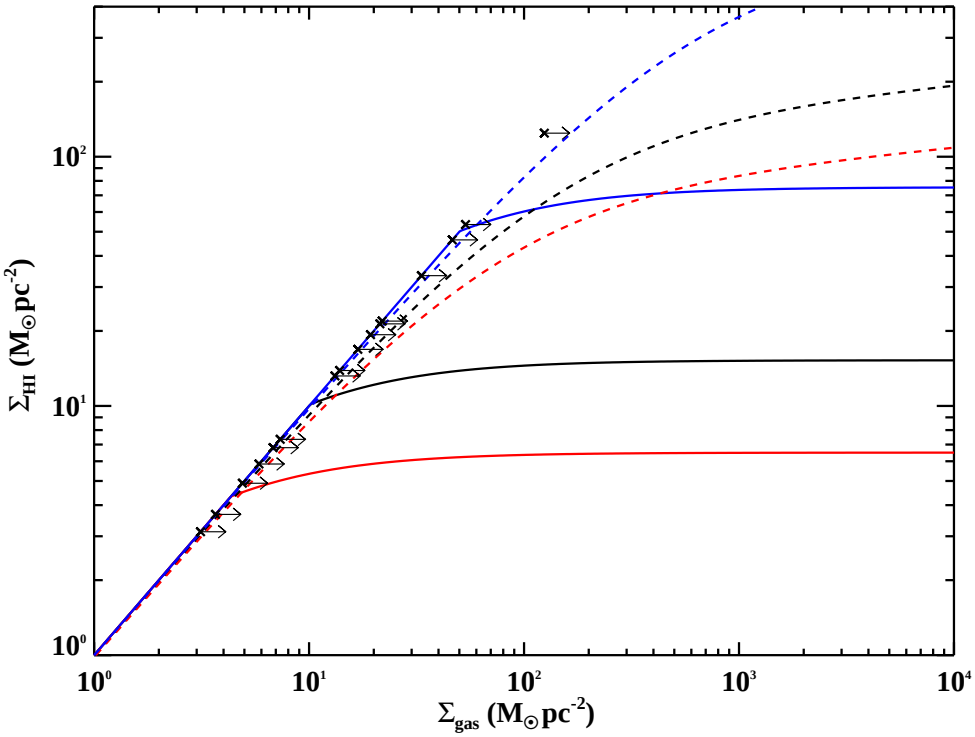

Figure 6.5: Comparison of the models with data in the low-resolution sample. Solid lines are for the KMT formalism, for the maximum ($Z' = 0.58$ red) central ($Z' = 0.20$ black) and minimum ($Z' = 0.03$ blue) observed metallicity. Dashed lines are for the BR model, for the maximum ($\log \rho'_{\rm star} = -1.45$ red), central ($\log \rho'_{\rm star} = -2.00$ black), and minimum ($\log \rho_{\rm star} = -3.18$ blue) stellar density in the sample. Lower limits on the total column density are for $\Sigma_{\rm gas} = \Sigma_{\rm HI}$. At low resolution, stellar densities drop by orders of magnitude and both models appear to be consistent with the data.

stellar density. Lower limits on the total gas column density are computed for $\Sigma_{\rm gas} = \Sigma_{\rm HI}$ (see Appendix 6.8.3 for a version of this figure that includes upper limits).

A comparison of Figures 6.5 and 6.4 reveals that the BR model predicts for the galaxy as a whole a much higher H I surface density, compared with the predictions for regions smaller than 100 pc. By going from the high-resolution sample to the low-resolution one, we lose the ability to analyse the local structure of the ISM, and are limited to average quantities which dilute the density contrasts in both gas and stars over many GMC complexes and aggregations. As a result, $\Sigma_{\rm HI}$ and $\rho_{\rm star}$ are lowered by one and two orders of magnitude respectively. This behaviour is reflected in the BR model as a decrease in the pressure by an order of magnitude, which now guarantees an overall agreement between data and the model. Conversely, the beam-smearing is accounted for in the KMT model by the clumping factor, here assumed to be $c \sim 5$. Hence, despite the different spatial scales, the KMT formalism is able to account for the mean observed $\Sigma_{\rm HI}$.

6.5.2.2 A test with individual galaxies

Although indicative of a general trend, Figure 6.5 does not allow a comparison of individual galaxies with models. This is particularly relevant for the KMT formalism that predicts a saturation in the H I column density as a function of the metallicity. To gain additional insight, we present in Figure 6.6 the predicted H I surface density ($\Sigma_{\rm HI,mod}$) against the observed one ($\Sigma_{\rm HI,obs}$) for individual galaxies. For the KMT model, we compute $\Sigma_{\rm HI,mod}$ using observed metallicities (red crosses). Since the total gas column density is unknown, we compute for each galaxy a range of $\Sigma_{\rm HI,mod}$ (shown with solid red lines) using upper and lower limits on the observed total gas surface density. Lower limits are derived assuming $\Sigma_{\rm gas} = \Sigma_{\rm HI,obs}$, while upper limits arise self-consistently from the H I saturation column density, naturally provided by the model. For the BR model, we instead compute only lower limits on $\Sigma_{\rm HI,mod}$ with stellar

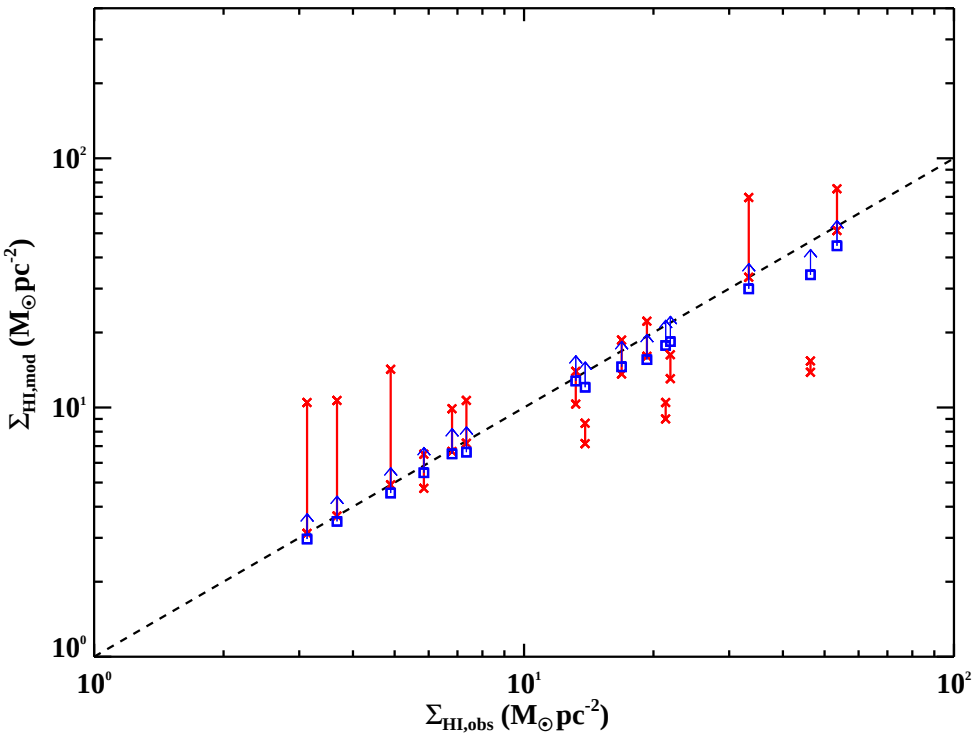

Figure 6.6: Predicted H I surface densities ($\Sigma_{\rm HI,mod}$) computed either with observed metallicity using the KMT model (red crosses) or with stellar density using the BR model (blue squares) are shown as a function of the observed values ($\Sigma_{\rm HI,obs}$) in individual galaxies. For the KMT model, lines connect lower and upper limits on $\Sigma_{\rm HI,mod}$, derived assuming proper limits on $\Sigma_{\rm gas}$ from the saturation threshold. The BR model is consistent with observations in individual galaxies, even at low metallicity. Similarly, the KMT formalism can account for most of the galaxies, but fail to reproduce observations in some cases.

densities (blue squares), assuming $\Sigma_{\rm gas} = \Sigma_{\rm HI,obs}$ (upper limits derived for $\Sigma_{\rm HI,mod}$ are presented in Appendix 6.8.3).

From Figure 6.6, the asymptotic behaviour of $\Sigma_{\rm HI}$ in the BR model prevents a tighter constraint on $\Sigma_{\rm gas}$. In general, there is good agreement between observations and model predictions for all the galaxies, despite the low mean metallicity of this sample. Conversely, the KMT model allows a narrower interval of H I surface density because of the well defined atomic hydrogen saturation. Therefore Figure 6.6 provides a more severe test of the KMT formalism

that nevertheless reproduces correctly most of the observations. Although for 4/16 galaxies the KMT predictions are inconsistent with the data, there seems to be no peculiar reason for the failure of the model for these objects (see Section 6.5.2.3).

For many objects the agreement between models and observations occurs close to the lower limits on $\Sigma_{\rm HI,mod}$, i.e. when $\Sigma_{\rm gas} = \Sigma_{\rm HI}$. We stress that this is not an obvious outcome of the assumption made on the total gas surface density, since high stellar densities in the BR model or high metallicity in the KMT model would imply a high molecular fraction irrespectively of $\Sigma_{\rm HI,obs}$. If we change perspective for a moment and assume that models are a reliable description of the molecular hydrogen content, the observed trend suggests that low-metallicity galaxies are indeed H I rich. Perhaps, this is not surprising due to the reduced dust content at low metallicity and the consequent reduction of the shielding of molecular gas from the LW-band photons.

6.5.2.3 Systematic effects for the KMT model

Finally, in Figure 6.7 we explore whether the KMT model exhibits systematic effects within the range of values allowed by the low-resolution sample. For this purpose, we present the ratio $\Sigma_{\rm HI,obs}/\Sigma_{\rm HI,mod}$ for the KMT model (red crosses) as a function of the observed metallicity (top panel) and stellar density (bottom panel). As in Figure 6.6, we display with solid lines the full interval of $\Sigma_{\rm HI,mod}$. The lack of any evident trend either with metallicity or stellar density suggests that the KMT model is free from systematic effects. In particular, because the four deviant galaxies are not found at systematically high or low metallicity, the difficulties in assessing metallicity for the cold-phase gas could be the cause of such deviations. This hypothesis was touched upon in Section 6.5.1.1, where we comment on the possibility that metals in the cold gas can in some cases be several factors lower than that observed in the ionized gas. An example of the importance of a correct metallicity determination is provided in Figure 6.8, where we repeat the comparison between the KMT model and observations in individual galaxies after

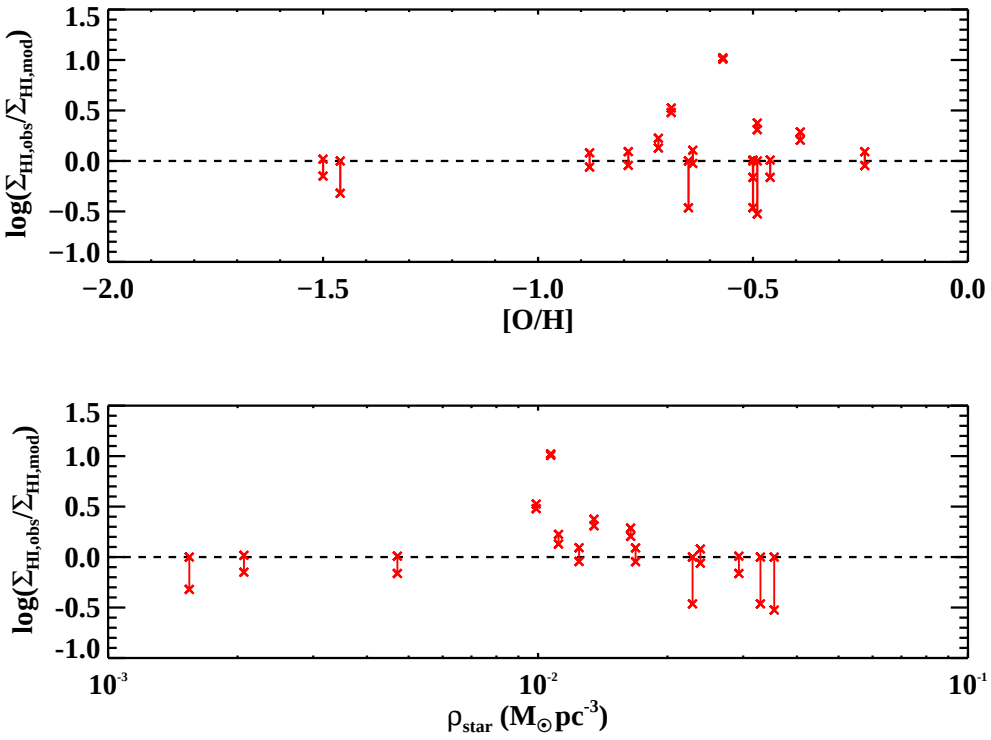

Figure 6.7: Ratio of the observed and predicted H I surface densities in the KMT model (red crosses) as a function of metallicity (upper panel) and stellar density (lower panel). The two dashed horizontal lines indicate the perfect agreement between data and model. There are no evident systematic trends in the discrepancies between data and the KMT formalism.

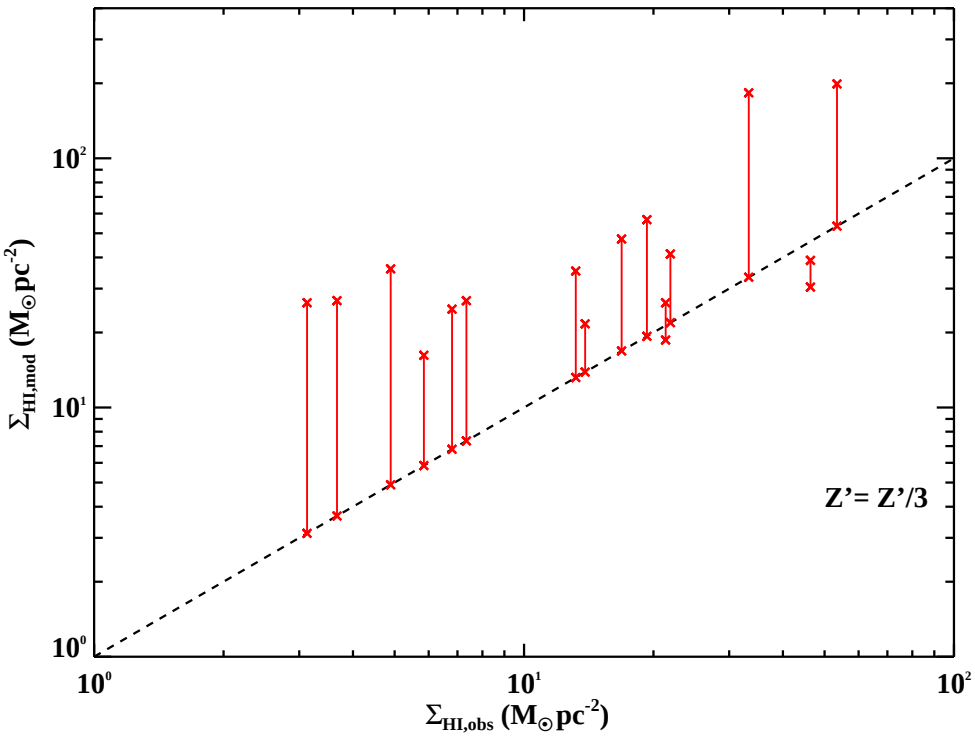

Figure 6.8: Comparison between the observed H I surface density and the value predicted by the KMT model, as in Figure 6.6, but for an arbitrarily lower metallicity ($Z' \equiv Z'/3$). A possible cause for the observed discrepancy data-model is a lower metallicity in the cold phase than the one observed in the ionized gas.

we arbitrarily redefine $Z' \equiv Z'/3$. As expected, for lower metallicity, the H I saturation moves to higher atomic surface densities and the KMT model better reproduces the observations.

6.6 Discussion

The analysis presented in Sect. 6.5 indicates that extrapolations of the BR model based on pressure overpredict the molecular fraction on small spatial scales (< 100 pc), while this formalism recovers the observed values on larger ones. Such a failure of the BR model on small scales was predicted by BR06, who point out that the model does not properly account

for the effects of gas self-gravity or local variations in the UV radiation field. Conversely, the KMT model based on gas and dust shielding is consistent with most of the observations both locally and on galactic scales, although it is more prone to observational uncertainties in the ISM structure (through the clumping factor) and the cold-phase metallicity. In this section, after a few comments on these points, we will focus on an additional result which emerges from our comparison: the molecular fraction in galaxies depends on the gas column density and metallicity, while it does not respond to local variations in pressure from enhancements in the stellar density.

6.6.1 The effect of self-gravity at small scales

The problem of gas-self gravity is that it introduces an additional contribution to the force balance on scales typical for GMCs. In this case, pressure equilibrium with the external ISM is no longer a requirement for local stability and the empirical power law in equation (6.7) may break down. However, since self-gravity enhances the internal pressure compared to the ambient pressure in equation (6.5), one would expect locally even higher molecular fractions than those predicted by an extrapolation of the BR model. This goes in the opposite direction of our results, since the observed molecular fraction is already overestimated. Hence, a different explanation must be invoked for the data-model discrepancy in Figure 6.3.

6.6.2 The effect of the radiation field

A reason for the high $\Sigma_{\rm HI}$ observed on small spatial scales is related to the intensity of the UV radiation field. In fact, regions which actively form stars probably have an enhanced UV radiation field compared to the mean galactic value. Since the BR formalism does not explicitly contain a dependence on the UV radiation field intensity, it is reasonable to expect discrepancies with observations. In contrast, the KMT model attempts to explicitly account for local variation

in j, by considering how such variations affect conditions in the atomic ISM. This makes it more flexible than the BR model in scaling to environments where conditions vary greatly from those averaged over the entire galaxy.

We stress here that the BR model is not completely independent of the radiation field, but simply does not account for a variation in j. Assuming the scaling relation $f_{\rm H_2} \sim P^{2.2} j^{-1}$ (Elmegreen 1993), the BR formalism is commonly considered valid only when variations in pressure are much greater than those in j, allowing us to neglect the latter. Being empirically based, the BR model contains information on a mean j, common for nearby spirals. Therefore this model describes the molecular content as if it were only regulated by pressure[5]. Hence, it can be applied to describe the molecular fraction only on scales large enough such that variations in the local UV intensity are averaged over many complexes and, eventually, j approaches a mean macroscopic value similar to that found in the galaxies used to fit the BR model.

This idea is quantitatively supported by recent numerical simulations. The molecular fraction of the galaxies simulated by Robertson & Kravtsov (2008) is consistent with the observed $R_{\rm H_2} \sim P^{0.92}$ only when the effects of the UV radiation field are taken into account. In fact, when neglecting the radiation field, a much shallower dependence $R_{\rm H_2} \sim P^{0.4}$ is found. In their discussion, the observed power-law index $\alpha \sim 0.9$ results from the combined effects of the hydrostatic pressure and the radiation field. Starting from $f_{\rm H_2} \propto P^{2.2} j^{-1}$ (Elmegreen 1993), assuming a Kennicutt-Schmidt law in which $j \sim \Sigma_{\rm sfr} \sim \Sigma_{\rm gas}^n$, and under the hypothesis that the stellar surface density is related to the gas surface density via a star formation efficiency $\Sigma_{\rm star} \sim \Sigma_{\rm gas}^\beta$, $R_{\rm H_2} = P^\alpha$ requires that (Robertson & Kravtsov 2008)

$$\alpha = 2.2 \frac{(1+\beta/2) - n}{1+\beta/2} \,. \tag{6.12}$$

[5]Incidentally, this exact statement can be found in BR06.

In the simulated galaxies, different star formation laws and efficiencies (mostly dependent on the galaxy mass) conspire to reproduce indexes close to the observed $\alpha \sim 0.92$, in support of the idea that a mean value for j is implicitly included in the empirical fit at the basis of the BR model.

6.6.3 The effect of stellar density

So far, we have discussed why the BR model extrapolated to small spatial scales is unable to predict the observed H I surface density due to its reliance on a fixed "typical" j. Galaxies in our sample are selected to be metal poor, but some of them have a higher SFR (median ~ 0.6 $M_\odot$ yr^{-1}) and higher specific star formation rate (SSFR; median $\sim 10^{-9}$ yr^{-1}) than observed in nearby spirals[6] ($< 10^{-10}$ yr^{-1}; e.g. Bothwell et al. 2009). If the UV intensity were the only quantity responsible for the disagreement between observations and the BR model below 100 pc, we would also expect some discrepancies on larger scales for galaxies with enhanced star formation. However, such discrepancies are not observed. As discussed, on larger scales, both the KMT and the BR models are able to reproduce observations, despite the different assumptions behind their predictions.

We argue that there is an additional reason for the observed discrepancy between data and the extrapolation of the BR model below 100 pc. While both models depend to first order on the gas column density, our analysis favours a model in which the local molecular fraction in the ISM depends to second order on metallicity rather than density as in the BR model. To illustrate the arguments in support of this hypothesis, in Figure 6.9 we show the atomic gas surface density predicted by the BR model (black lines) as a function of the stellar density. Different curves are for different values of the total gas surface density ($\Sigma'_{\rm gas} = 10$ solid line, $\Sigma'_{\rm gas} = 15$ dotted

[6]We recall our choice of the maximum SFR among the values available in the literature to conservatively obtain an upper limit on the molecular emission. For this reason, our median SFR is biased towards high values. In any case, a SFR ~ 0.3 $M_\odot$ yr^{-1}, typical of BCDs (Hopkins et al. 2002), would correspond to a SSFR 5×10^{-10} yr^{-1} at the median mass $M_{\rm star} = 6 \times 10^8$ $M_\odot$.

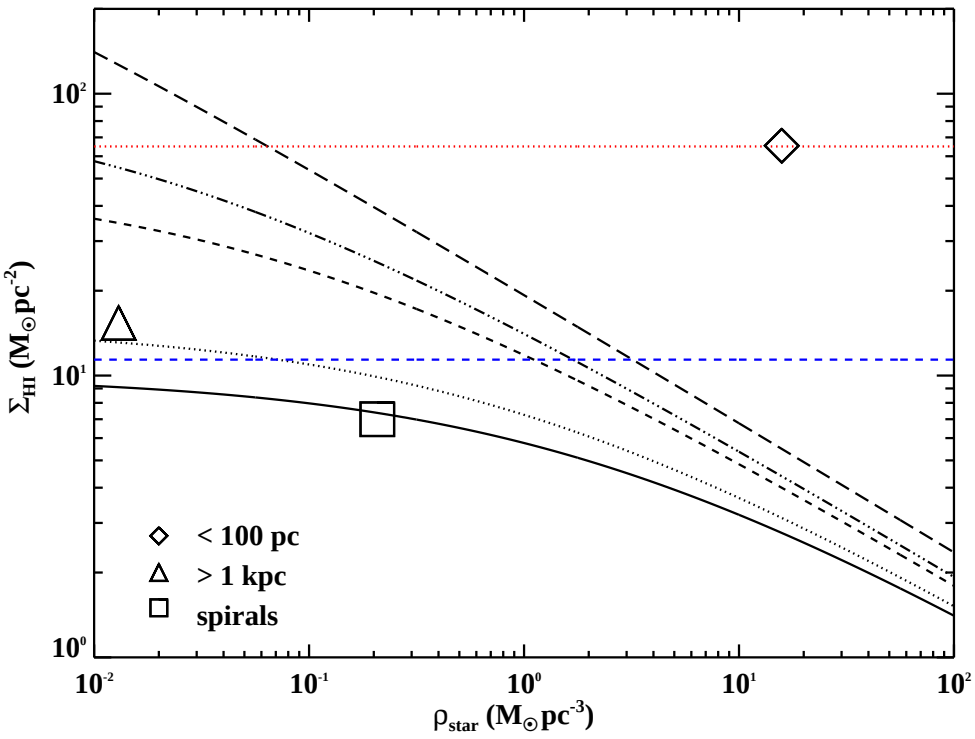

Figure 6.9: Atomic gas surface density predicted by the BR model (black lines) as a function of the stellar density. Different curves are for different values of the total gas surface density ($\Sigma'_{\rm gas} = 10$ solid line, $\Sigma'_{\rm gas} = 15$ dotted line, $\Sigma'_{\rm gas} = 50$ dashed line, $\Sigma'_{\rm gas} = 10^2$ dash-dotted line, and $\Sigma'_{\rm gas} = 10^3$ long-dashed line). The two horizontal lines are for the KMT model at the mean metallicity in the high-resolution sample (red-dotted line) and in the low-resolution sample (blue-dashed line). The open diamond and the open triangle represent the median values of stellar density and H I surface density in the high-resolution and low-resolution samples, respectively. The open square represents a typical stellar density and H I surface density for local spiral galaxies, from a median of the values in table 1 of BR04, assuming a stellar disk height $h = 300$ pc. Our analysis favours a model in which the molecular fraction in the ISM depends to second order on metallicity rather than stellar density.

line, $\Sigma_{\rm gas}^{'} = 50$ dashed line, $\Sigma_{\rm gas}^{'} = 10^2$ dash-dotted line, and $\Sigma_{\rm gas}^{'} = 10^3$ long-dashed line). Superimposed, there are three data points. The open diamond and the open triangle represent the median values of stellar density and H I surface density for the high-resolution ($\Sigma_{\rm HI}^{'} = 65$, $\rho_{\rm star}^{'} = 16$) and low-resolution ($\Sigma_{\rm HI}^{'} = 15$, $\rho_{\rm star}^{'} = 0.01$) samples, respectively. The open square represents a typical stellar density and H I surface density ($\Sigma_{\rm HI}^{'} = 7$, $\rho_{\rm star}^{'} = 0.2$) for local spiral galaxies, from a median of the values in table 1 of BR04, assuming a stellar disk height $h = 300$ pc. Finally, the two horizontal lines are for the KMT model at the mean metallicity in the high-resolution sample (red-dotted line, $Z^{'} = 0.12$) and in the low-resolution sample (blue-dashed line, $Z^{'} = 0.20$).

Recalling that for the BR model $P \sim \Sigma_{\rm gas} v_{\rm gas} \sqrt{\rho_{\rm star}}$, we see from Figure 6.9 that the stellar density does not provide a major contribution to the variation in pressure when $\rho_{\rm star} \lesssim 0.5$, typical for large galactic regions (open triangle and square). In this regime, the predicted $\Sigma_{\rm HI}$ from the BR model (black lines) is mostly dependent on varaition in the total gas column density alone. In fact, for a constant velocity dispersion, the pressure model becomes equivalent to models based on gas shielding. This reconciles on theoretical grounds the equivalence observed in local spirals between the BR and KMT (blue horizontal dashed line) models. At local scales, i.e. moving towards higher stellar density, $\rho_{\rm star}$ provides an important contribution to the total pressure and the extrapolation of the BR model predicts high molecular gas fractions, as seen in Figure 6.9. The expected $\Sigma_{\rm HI}$ drops accordingly, following a trend that is in disagreement with observations (open diamond). Conversely, the KMT model is insensitive to $\rho_{\rm star}$ and accounts only for a variation of the gas column density and metallicity. In this case, the predicted $\Sigma_{\rm HI}$ (red horizontal dotted line) moves towards higher values in agreement with observations.

From this behaviour, we conclude that stellar density is not a relevant quantity in determining the *local* molecular fraction. Furthermore, because the KMT formalism recovers the high observed H I column density for low metallicity, dust and metals have to be important

in shaping the molecular content of the ISM. This is further supported by Gnedin & Kravtsov (2010) who show with simulations how the observed star formation rate, which mostly reflects the molecular gas fraction, depends on the metallicity[7]. It follows that the observed dependence on the midplane pressure is only an empirical manifestation of the physics which actually regulates the molecular fraction, i.e. the effects of the UV radiation field and the gas and dust shielding.

6.6.4 Fixed stellar density for molecular transitions

A final consideration regards the observational evidence that the transition from atomic to molecular gas occurs at a fixed stellar density with a small variance among different galaxies (BR04). This result favors hydrostatic pressure models because if the atomic-to-molecular transition were independent of stellar surface density, there would be much more scatter in the stellar surface density than is observed. However, this empirical relation can also be qualitatively explained within a formalism based on UV radiation shielding. In fact, if the bulk of the star formation takes place in molecule dominated regions, the build-up of the stellar disks eventually will follow the molecular gas distribution, either directly or via a star formation efficiency (see Robertson & Kravtsov 2008, Gnedin & Kravtsov 2010). Since the transition from molecular to atomic hydrogen occurs at a somewhat well-defined gas column density (KMT09; WB02; Bigiel et al. 2008), it is plausible to expect a constant surface stellar density at the transition radius. While this picture would not apply to a scenario in which galaxies grow via subsequent (dry) mergers, recent hydrodynamical simulations (e.g. Brooks et al. 2009) support a model in which stars in disks form from in-situ star formation from smoothly accreted cold or shock-heated gas.

[7] However, a detailed analysis of star forming regions in BCDs by Hirashita & Hunt (2004) suggests that other parameters such as gas density, size, and geometry play a role in determining the local SFR.

6.7 Summary and conclusion

With the aim of understanding whether the principal factor that regulates the formation of molecular gas in galaxies is the midplane hydrostatic pressure or shielding from UV radiation by gas and dust, we compared a pressure model (BR model; Wong & Blitz 2002, Blitz & Rosolowsky 2004; 2006) and a model based on UV photodissociation (KMT model; Krumholz et al. 2008; 2009a, McKee & Krumholz 2010) against observations of atomic hydrogen and stellar density in nearby metal-poor dwarf galaxies. Due to their low metallicity and high stellar densities, these galaxies are suitable to disentangle the two models, otherwise degenerate in local spirals because of their proportionality on the gas column density.

Our principal findings can be summarized as it follows.

- On spatial scales below 100 pc, we find that an extrapolation of the BR model (formally applicable above $\sim$ 400 pc) significantly underpredicts the observed atomic gas column densities. Conversely, observations do not disfavour predictions from the KMT model, which correctly reproduces the high H I gas surface densities commonly found at low metallicities.

- Over larger spatial scales, with the observed and predicted H I surface density integrated over the entire galaxy, we find that both models are able to reproduce observations.

- Combining our results with numerical simulations of the molecular formation in the galaxies ISM (Elmegreen 1989, Robertson & Kravtsov 2008) which indicates how the UV radiation field (j) plays an essential role in shaping the molecular fraction, we infer that the discrepancy between the BR model and observations on *local scales* is due partially to the model's implicit reliance on an average j, which breaks down at small scales. In contrast, the KMT model properly handles this effect.

- Since on scales $\sim$ 1 kpc the BR model agrees with observations despite the low metallicity and high specific SFR in our sample, we infer that the discrepancy between pressure models and observations below 100 pc also arises from their dependence on stellar density. An increase in stellar density corresponds to an increase in the hydrostatic pressure which should, in the BR model, reduce the atomic gas fraction. No such trend is seen in the observations.

- If we drop the dependence on the stellar density, the pressure model reduces to a function of the total gas column density and becomes equivalent to the KMT model, for a fixed velocity dispersion and metallicity. This provides a theoretical explanation for the observed agreement of the two models in local spirals.

In conclusion, our analysis supports the idea that the local molecular fraction is determined by the amount of dust and gas which can shield H_2 from the UV radiation in the Lyman-Werner band. Pressure models are only an empirical manifestation of the ISM properties, with the stellar density not directly related to the H_2 formation. Although they are useful tools to characterize the molecular fraction on large scales, obviating the problem of determining the clumpy structure of the ISM or the metallicity in the cold gas as required by models based on shielding from UV radiation, pressure models should be applied carefully in environments that differ from the ones used in their derivation. These limitations become relevant in simulations and semi-analytic models, especially to describe high-redshift galaxies. Furthermore, a correct understanding of the physical processes in the ISM is crucial for the interpretation of observations, an aspect that will become particularly relevant once upcoming facilities such as ALMA will produce high-resolution maps of the ISM at high redshifts.

Combining our analysis with both theoretical and observational efforts aimed at the description of the ISM characteristics and the SFR in galaxies, what emerges is a picture in

which macroscopic (hence on galactic scales) properties are regulated by microphysical processes. Specifically, the physics that controls the atomic to molecular transition regulates (and is regulated by) the SFR, which sets the UV radiation field intensity. The ongoing star formation is then responsible for increasing the ISM metallicity and building new stars, reducing and polluting at the same time the primordial gas content. Without considering violent processes more common in the early universe or in clusters, this chain of events can be responsible for a self-regulated gas consumption and the formation of stellar populations. Future and ongoing surveys of galaxies with low-metallicity, active star formation and high gas fraction (e.g., LITTLE THINGS; Hunter et al.) will soon provide multifrequency observations suitable to test in more detail the progress that has been made on a theoretical basis to understand the process of star formation in galaxies.

6.8 Appendix

6.8.1 Notes on individual galaxies

I Zw 18

The main body of I Zw 18 consists of two main clusters, the north-west (NW) and the south-east (SE) components with an angular separation of $\sim 6''$. A third system, known as "Zwicky's flare" or "C" component, lies about $22''$ to the northwest of the NW cluster. H I maps are available from van Zee et al. (1998a), together with the *HST*/WFPC2 F814W image ($0\farcs045$ resolution). The H I peaks close to the fainter SE cluster, rather than to the NW where the stellar density is higher. Lyα observations with *HST*/GHRS by Kunth et al. (1994) ($2''\times 2''$ beam) are also available for the NW cloud and have a better resolution than the VLA map. At the assumed distance of 13 Mpc (Izotov & Thuan 2004b), $1'' = 63$ pc. Stellar masses of

the two massive clusters in I Zw 18 are not published. Hence, multiband integrated photometry of the two star clusters in I Zw 18 is taken from Hunt et al. (2003), together with cluster ages as modeled by them. Sizes of the clusters are measured from fitting Gaussians to the surface brightness profiles (not previously published).

The SE cluster has an age of 10 Myr and, near where the distribution peaks (see van Zee et al. 1998a), $M_{\rm K} = -12.4$ (in a $2''$ aperture). The lowest-metallicity SB99 models ($Z' = 0.001$) give $M_{\rm K} = -12.4$, which implies a stellar mass of $2.3 \times 10^5\,{\rm M}_\odot$. With a K-band luminosity of $1.91 \times 10^6 {\rm L}_\odot$, this would give a $(M/L)_{\rm K} = 0.12$, as inferred from the SB99 models. The Bell & de Jong (2001) predictions give $(M/L)_{\rm K} = 0.17$, on the basis of $V - K$ and $(M/L)_{\rm K} = 0.09$, from $V - J$. Hence, the SB99 value of 0.12 is roughly consistent. We therefore adopt the value of $2.3 \times 10^5\,{\rm M}_\odot$ for the stellar mass of the SE cluster. We can check the inferred stellar mass by inspecting the K-band surface brightness at the SE peak (see Figure 5 in Hunt et al. 2003), $\mu_{\rm K} = 19.2$ mag arcsec^{-2}. This gives $\Sigma_{\rm K} = 185.3$ L$_\odot$ pc^{-2} and, assuming $(M/L)_{\rm K} = 0.12$ from SB99, we would have $\Sigma_{\rm star} = 22.2\,{\rm M}_\odot{\rm pc}^{-2}$. This is in good agreement with the value of $\Sigma_{\rm star} = 23.3\,{\rm M}_\odot{\rm pc}^{-2}$, inferred from the absolute luminosity (see above) and the measured radius of 56 pc.

The NW cluster has an age of 3 Myr and $M_{\rm K} = -13.25$ (in a $2''$ aperture). The lowest-metallicity SB99 models give $M_{\rm K} = -16.2$, which is rather uncertain because of the rapid increase in luminosity at about 3 Myr when the most massive stars start evolving off of the main sequence, a phase which is not correctly described in models (Origlia et al. 1999). In fact, the observed $V - H$ color of 0.29 is predicted by SB99 to occur at $\sim$10 Myr, not at 3 Myr, which is the best-fit photometric age. In any case, the inferred mass from this model is $6.6 \times 10^4\,{\rm M}_\odot$. With a K-band luminosity of 1.91×10^6 L$_\odot$, this would give a $(M/L)_{\rm K} = 0.016$. The same exercise repeated for the V band, with $M_{\rm V} = -12.86$ and the SB99 prediction of $M_{\rm V} = -15.7$, give an inferred mass of $7.3 \times 10^4\,{\rm M}_\odot$, and $(M/L)_{\rm V} = 0.006$. These M/L values are quite

low, roughly 6 times smaller than those predicted by Bell & de Jong (2001) from the observed colors of I Zw 18. We therefore use the latter M/L ratio. With $V-K=0.38$, $V-H=0.29$, and $V-I=-0.04$, we estimate $(M/L)_{\rm K}=0.11$, 0.10, and 0.09, respectively. Therefore, adopting 0.10, we derive a stellar mass of $4.2\times10^5\,{\rm M}_\odot$. Repeating the calculations for V band, we find $(M/L)_{\rm V}=0.039$, 0.034, and 0.033, respectively. Adopting 0.033, we would derive a similar stellar mass of $3.9\times10^5\,{\rm M}_\odot$. Again, the K-band surface brightness of the NW cluster ($\mu_{\rm K}=18.3$ mag arcsec^{-2}, $0\farcs5$ resolution) gives a similar result. We find $\Sigma_{\rm K}=536\ {\rm L}_\odot\ {\rm pc}^{-2}$, and, with $(M/L)_{\rm K}=0.10$, becomes 53.6 M pc^{-2} . With a cluster radius of 56 pc ($0\farcs89$), this would correspond to a cluster mass of $5.3\times10^5\,{\rm M}_\odot$, about 1.3 times that inferred from the lower-resolution photometry. Hence, to obviate problems of resolution ($1''$ radius aperture or 63 pc, vs. a 56 pc radius measured from the *HST* image), we adopt the mean of these two measurements for the stellar mass of the NW cluster, namely $4.7\times10^5\,{\rm M}_\odot$.

SBS 0335−052 E

SBS 0335−052 E hosts six SSCs, with most of the star formation activity centered on the two brightest ones to the southeast. The H I distribution is published in Ekta et al. (2009), and the *HST*/ACS F555M image ($0\farcs050$ resolution), was published by Reines et al. (2008). The H I map is of relatively low resolution ($\sim3\farcs4$) and does not resolve the six SSCs individually since they are distributed (end-to-end) over roughly $2\farcs6$. Lyα observations with *HST*/GHRS by Thuan & Izotov (1997) ($2''\times2''$ beam) are also available. In our analysis, we use this column density, being at better resolution than the one derived from H I emission map. At the assumed distance of 53.7 Mpc, $1''=260.3$ pc. Stellar masses for individual clusters have been derived by Reines et al. (2008) by fitting the optical and UV spectral energy distributions, and we adopt these masses here. Comparison with masses inferred from K-band is unfruitful since nebular and ionized gas contamination make this estimate highly uncertain. We measure the size of the

clusters by fitting two-dimensional Gaussians. They are unresolved at the *HST*/ACS resolution of 0.″050, but since they have the same size to within 13%, we assume the average radius of 18.2 pc. Therefore, the inferred mass densities result in lower limits.

Mrk 71

Mrk 71 (NGC 2363) is a complex of H II regions in a larger irregular galaxy, NGC 2366. There are two main knots of star-formation activity (see Drissen et al. 2000), denoted A and B. A low-resolution H I map (12.″5×11.″5) is available from Thuan et al. (2004), but at this resolution we are unable to distinguish the two main clusters which are 5″ apart. At the assumed distance of 3.44 Mpc (derived from Cepheids, Tolstoy et al. 1995), 1″ = 16.7 pc. Stellar masses of the two starburst knots in Mrk 71 are not published. Hence, V-band photometry was taken from Drissen et al. (2000), and I-band from Thuan & Izotov (2005), together with cluster ages as modeled by Drissen et al. (2000). As for I Zw 18, sizes of the clusters were measured from fitting 1D Gaussians to the surface brightness profiles.

The knot A has $V = 17.3$ mag and, after correcting for $A_{\rm V} = 0.3$ mag, we derive an absolute magnitude $M_{\rm V} = -10.4$. At an age of 3 Myr, SB99 models (at Z=0.004) predict $M_{\rm V} = -15.2$. We would thus infer a stellar mass of $1.2 \times 10^4\,{\rm M}_\odot$, and an implied $(M/L)_{\rm V}$ ratio of 0.012. The Bell & de Jong (2001) predictions give $(M/L)_{\rm V} = 0.010$, on the basis of stellar $V-K = -0.42$, as modeled by Noeske et al. (2000), and $(M/L)_{\rm V} = 0.10$, from stellar $B-V = -0.19$ (also as in Noeske et al. 2000). The latter value from $B-V$ is a factor of 10 higher than the former from $V-K$, and highly inconsistent with the SB99 value of 0.01. The I-band photometry of knot A from Thuan & Izotov (2005) gives a similar inconsistency. With $I = 17.97$ mag, and a corresponding absolute magnitude of 9.71, we would infer a stellar luminosity of $3.43 \times 10^5\ {\rm L}_\odot$. With Bell & de Jong (2001) $(M/L)_{\rm I}$ values of 0.025 (from $V-K$) and 0.15 (from $B-V$), we would derive stellar masses of 8.6×10^3 and $5.1 \times 10^4\,{\rm M}_\odot$, respectively.

Since three values are roughly consistent ($\sim 10^4\,M_\odot$), we adopt $1.2 \times 10^4\,M_\odot$ as the mass for knot A.

Knot B is slightly older than knot A (4 Myr) consistent with its Wolf-Rayet stars and strong stellar winds as inferred from P Cygni-like profiles in the UV (Drissen et al. 2000). It is also slightly fainter with $V = 18.05$ (after correcting for $A_V = 0.3$ mag), corresponding to an absolute magnitude $M_V = -9.63$. SB99 models (at 4 Myr) predict $M_V = -15.3$, which would give a stellar mass of $5.4 \times 10^3\,M_\odot$, and an implied $(M/L)_V$ ratio of 0.005. Again, we derive M/L ratios as a function of color from Bell & de Jong (2001), and obtain $(M/L)_V = 0.06$ from stellar $V - K = 0.67$ and $(M/L)_V = 0.15$, from stellar $B - V = -0.07$ (Noeske et al. 2000). With a stellar V-band luminosity of $6.1 \times 10^5\ L_\odot$, we would infer a stellar mass of $3.6 \times 10^4\,M_\odot$ with $(M/L)_V = 0.06$, and $9.2 \times 10^4\,M_\odot$ with $(M/L)_V = 0.15$. Both masses are larger than those inferred for knot A, inconsistently with the observation of Drissen et al. (2000) that knot B contains only ∼6% of the ionizing photons necessary to power the entire H II region. Nevertheless, knot A is supposedly enshrouded in dust (Drissen et al. 2000), so the situation is unclear. The I-band photometry of knot B from Thuan & Izotov (2005) is not edifying. With $I = 18.92$ mag, and a corresponding absolute magnitude of −8.76, we would infer a stellar luminosity of $1.43 \times 10^5\ L_\odot$. With Bell & de Jong (2001) $(M/L)_I$ values of 0.10 (from $V-K$) and 0.20 (from $B-V$), we would derive stellar masses of $1.4 \times 10^4\,M_\odot$ and $2.9 \times 10^4\,M_\odot$, respectively. These values are all greater than the mass inferred for knot A, even though knot B is reputed to be intrinsically ∼16 times fainter (see above, and Drissen et al. 2000). For this reason, we adopt the SB99 value of $5.4 \times 10^3\,M_\odot$ as the mass for knot B.

UM 462

UM 462 hosts six SSCs, with 9″ separation from end to end. The H I distribution is available in the literature from van Zee et al. (1998b), together with the ground-based

ESO/SOFI Ks image (0.″28 pixels, 0.″8-1″seeing) from Vanzi (2003). The resolution of the H I map (6.″6×5.″2) is just barely sufficient to distinguish the two clusters. At the assumed distance of 13.5 Mpc, $1'' = 65.4$ pc. Stellar masses of the six SSCs in UM 462 have been derived by Vanzi (2003) by estimating the age from the Hα equivalent width, then comparing SB99 models at that age with the Hα luminosity of the cluster after correcting for extinction. These estimates differ from the other SB99 comparisons described for previous galaxies because they extend to a lower lower-mass limit, 0.1 $M_\odot$ rather than 1 $M_\odot$. We thus consider a range of possible masses given by the values published by Vanzi (2003) and what we infer from comparing the Ks-band luminosities of the individual clusters with SB99 predictions as above (at the published age). On the basis of the observed $V-Ks$ and $V-J$ colors (Vanzi et al. 2002, Vanzi 2003), we derive $(M/L)_{\rm K} = 0.05$ according to Bell & de Jong (2001). A comparison of these numbers with the values provided by Vanzi (2003) reveals that typical uncertainty on the mass calculations are roughly a factor of 2 or less. Sizes are measured by Vanzi et al. (2002), but the clusters are unresolved at the ground-based resolution. Hence, the mass densities are formally lower limits.

II Zw 40

II Zw 40 is a cometary BCD with two tails. The main star-formation activity is occurring at the "head of the comet", namely in two knots in a north-south orientation, separated vertically by 1.″5. The upper knot, dubbed "A" by Vanzi et al. (2008), is elongated along roughly an east-west direction and contains the rising-spectrum thermal radio sources found by Beck et al. (2002). The lower B knot is round, fainter than A, and apparently does not host any compact radio sources. The H I distribution is published in van Zee et al. (1998b) and *HST*/F814W images (0.″025 pixels) are also available. The resolution of the H I map, 5.″7×4.″8, does not distinguish the two knots seen at *HST* resolution. At the assumed distance of 10.3 Mpc, $1'' = 49.9$ pc. II Zw 40 is located sufficiently near the plane of the Milky Way that the

foreground extinction is quite high, $A_{\rm V} = 2.7$ mag. Stellar masses of the two clusters in II Zw 40 have been derived by Vanzi et al. (2008), but with a Kroupa IMF and using Brγ luminosity (see also UM 462) in a 15″ aperture. Moreover, a subtraction procedure was applied that made assumptions about the Brγ flux of region A in the vicinity of knot B. Hence, to test these values for stellar masses, we recompute them from continuum measurements, using similar procedures to those used for the previous galaxies.

Ages for the clusters are taken from Vanzi et al. (2008), who compared observed Brγ equivalent widths to SB99 model predictions. Photometry in the *HST* F555W, F814W, and F160W passbands is performed by fitting a 2D Gaussian to each knot. This determination is roughly consistent with aperture photometry in a 1″ aperture. Correcting the observed magnitudes for the high extinction as in Vanzi et al. (2008) from the NIR hydrogen recombination lines ($A_{\rm V} = 4.0$ mag for knot A, $A_{\rm V} = 4.9$ mag for knot B), we have 14.62, 15.80, 13.34 for F555W, F814W, and F160W, respectively (knot A), and 14.13, 14.58, and 13.31 for F555W, F814W, and F160W (knot B). Under the approximation F555W∼V , we derive an absolute magnitude of −15.4 for knot A, to be compared with the SB99 prediction of −15.2 (at 3 Myr), and −15.9 for knot B to be compared with −15.5 (at 7 Myr). The comparisons with SB99 give stellar masses of $1.2 \times 10^6\,{\rm M}_\odot$ and $1.5 \times 10^6\,{\rm M}_\odot$, for knots A and B, respectively (the inferred $(M/L)_{\rm V}$ are 0.01 and 0.007 for knots A and B). With the approximation F555W∼V , F814W∼I, and F160W∼H, $V - I = -1.18$ and -0.45, and $V - H = 1.31$ and 0.82 for knots A and B, respectively. Using Bell & de Jong (2001), we derive $(M/L)_{\rm V} = 0.0005$ and 0.004 from $V - I$, and $(M/L)_{\rm V} = 0.21$ and 0.09 from $V - H$.

These values differ substantially from those inferred from SB99. In particular, the value of 0.0005 is unrealistic, and difficult to reconcile with other galaxies and other M/L inferred for II Zw 40. Hence, we consider a range of stellar masses for knot A with $6.4 \times 10^5\,{\rm M}_\odot$ (for $(M/L)_{\rm V} = 0.005$) and $1.2 \times 10^6\,{\rm M}_\odot$ (SB99) as lower and upper limits. For knot B instead, we

assume a lower limit at $8.1 \times 10^5\,M_\odot$ (for $(M/L)_V = 0.004$) and an upper limit at $1.5 \times 10^6\,M_\odot$ (SB99). These are roughly consistent with the masses given by Vanzi et al. (2008) of $1.7 \times 10^6\,M_\odot$ for knot A and $1.3 \times 10^5\,M_\odot$ for knot B, as inferred from comparing SB99 predictions of Brγ emission over a $15''$ aperture (knot A) and $0\farcs75$ aperture (knot B). For the sizes of the two clusters, we fit a 1D Gaussians to the surface-brightness profiles in the F814W band and obtain 10.1 pc for knot A and 5.2 pc for knot B. The dimensions derived by fitting 2D Gaussians are smaller, namely 4.4 pc (A) and 3.7 pc (B), similar to the dimensions obtained by Vanzi et al. (2008) for the star cluster itself (rather than the more extended H II region emission).

NGC 5253

NGC 5253 is a nearby dwarf galaxy in the Centaurus group at a distance of 3.5 Mpc. Its morphology is peculiar; the outer isophotes resemble a dwarf elliptical, but over time NGC 5253 has been classified as a spiral, an elliptical, an S0, an irregular, and most recently, as an amorphous galaxy (Caldwell & Phillips 1989). A blue starburst dominates the central region, with a dust lane bisecting the main body along the minor axis. The central starburst comprises at least six SSCs, identified by Calzetti et al. (1997), who published *HST* multiband *HST*/WFPC2 images of the galaxy with a resolution of $0\farcs1$. The H I distribution is published by Kobulnicky & Skillman (2008), with a beam size of $9\farcs0 \times 7\farcs6$.

Calzetti et al. (1997) have measured the cluster ages by comparing colors and equivalent width of hydrogen recombination lines with SB99 predictions. The reddest cluster, NGC 5253-5, has an age of $\lesssim 2.5$ Myr, and dominates the infrared spectral energy distribution (Vanzi & Sauvage 2004); its visual extinction A_V is uncertain but could be as large as 35 mag (Calzetti et al. 1997), although is probably around 7-8 mag (Vanzi & Sauvage 2004). The brightest and bluest cluster, NGC 5253-4, is also quite young, $\sim$2.5 Myr, but the remaining SSCs are older, $\sim$10-50 Myr. Stellar masses of the SSCs have been inferred by Calzetti et al. (1997) and

Vanzi & Sauvage (2004), through comparison of the observed broadband luminosities to SB99 predictions, given the age of the cluster. Masses range from $\sim 10^4\,M_\odot$ (NGC 5253-4) to $10^6\,M_\odot$ (NGC 5253-5), and radii from 1.6 to 3.5 pc, as measured from *HST* images.

NGC 1140

NGC 1140 is an amorphous, irregular galaxy, and, like NGC 5253, has been reclassified over the course of time (Hunter et al. 1994b). Optically, it is dominated by a supergiant H II region encompassing $\sim 10^4$ OB stars, far exceeding the stellar content of the giant H II region, 30 Doradus, in the LMC (Hunter et al. 1994b). The H II region is powered by several SSCs (Hunter et al. 1994b, de Grijs et al. 2004), situated in a vertical strip, about 10″ in length. High-resolution *HST*/WFPC2 images exist for this galaxy (Hunter et al. 1994b), and H I maps have been published by Hunter et al. (1994a). The resolution of the H I map (16″×22″) is insufficient to distinguish the clusters; given this resolution, the peak $\Sigma_{\rm HI}$ is certainly underestimated.

Stellar masses have been determined for the SSCs in NGC 1140 by de Grijs et al. (2004), using a minimization technique which simultaneously estimates stellar ages and masses, metallicities and extinction using broadband fluxes. We prefer these masses to those measured by a virial technique (Moll et al. 2009), as the latter can be an order of magnitude larger, perhaps due to non-virial line widths. Sizes were measured by Moll et al. (2009) for the brightest clusters 1 and 6, and we have used these, after correcting for the different distance scale.

6.8.2 Stellar masses for the low-resolution sample

To compute stellar masses for the low-resolution sample of 16 objects, we acquired IRAC 3.6 and 4.5 μm data from the Spitzer archive for all galaxies in the sample (3.6 μm images are unavailable for two objects). Starting from the Basic Calibrated Data, we coadd frames using MOPEX, the image mosaicing and source-extraction package provided by the Spitzer Science

center (Makovoz & Marleau 2005). Pixels flagged by masks are ignored. Additional inconsistent pixel values are removed by means of the MOPEX outlier rejection algorithms, in particular the dual-outlier technique, together with the multiframe algorithm. We correct the frames for geometrical distortion and then project them onto a fiducial coordinate system with pixel sizes of $1\farcs2$, roughly equivalent to the original pixels. Standard linear interpolation is used for the mosaics. The noise levels in our MOPEX IRAC mosaics are comparable to or lower than those in the pipeline products. We then perform aperture photometry on the IRAC images with the IRAF photometry package `apphot` and applying appropriate unit conversion to compute integrated fluxes. The background level is determined by averaging several adjacent empty sky regions. Fluxes are computed with a curve-of-growth analysis at radii where the growth curve becomes asymptotically flat. From this analysis, we determined the half-light (effective) diameter and the size of the object as the point at which the growth curve achieves flatness. Our values for IRAC sizes are on average 1.24 times the geometric means of the optical dimensions (as reported in NED). Although the standard deviation is large (0.48), this could indicate that there is an evolved stellar population in the extended regions of the galaxies that is not seen at optical wavelengths. Alternatively, it could merely be an effect of surface brightness since the optical diameters are isophotal and, for a given optical surface brightness, the IR images could be deeper. In addition, the two methods of measuring sizes also differ (isophotal in the optical, and photometric in the IR). In fact, the approximation of a circular virtual aperture could contribute to the larger sizes measured with IRAC.

From the IRAC photometry, we derive stellar masses following Lee et al. (2006) by inferring K-band luminosities from IRAC [4.5] total magnitudes (with a color correction), and a color-dependent $(B-K)$ mass-to-light ratio. To increase the reliability of this procedure, we include and calibrate with a similar procedure the IRAC [3.6] magnitudes, and, where available, also incorporate K-band magnitudes from 2MASS. The average of these three values (one K

band, and two indirect K-band estimates) are used to calculate the $B-K$ color for the M/L ratio, and the K-band luminosity. These magnitudes and colors are reported in Table 6.3.

Table 6.3: Photometric quantities for the low resolution sample[a]

Name	B (mag)	K (mag)	K ([3.6])[b] (mag)	K ([4.5])[c] (mag)	$B-K$ (min.) (mag)	$B-K$ (mean) (mag)	$B-K$ (std) (mag)	$M_{star,K}$ (min.) ($M_\odot$)	$M_{star,K}$ (avg) ($M_\odot$)
Haro3	13.22	10.61	-	10.05	2.61	2.89	0.40	9.09	9.35
IIZw40	11.94	10.98	10.89	10.22	0.96	1.24	0.42	8.09	8.34
IZw18	15.86	15.24	14.59	14.45	0.62	1.10	0.42	6.51	6.88
Mrk209	15.09	12.50	-	12.68	2.40	2.49	0.12	7.29	7.36
Mrk33	13.39	10.42	10.09	10.13	2.96	3.17	0.18	9.59	9.79
Mrk71	11.44	-	11.74	10.90	-0.30	0.12	0.60	6.54	6.89
NGC1140	13.46	10.48	9.92	10.07	2.98	3.30	0.29	9.30	9.58
NGC1156	11.35	9.45	8.49	8.51	1.90	2.53	0.55	8.62	9.10
NGC1741	13.08	11.82	10.76	10.78	1.26	1.96	0.61	9.29	9.81
NGC2537	12.09	9.11	8.91	9.02	2.98	3.07	0.10	9.14	9.26
NGC4214	10.15	7.90	7.92	-	2.22	2.23	0.02	8.66	8.67
NGC5253	10.63	8.21	7.60	7.23	2.41	2.95	0.50	8.63	9.05
NGC7077	13.90	11.36	11.39	11.47	2.42	2.49	0.06	8.54	8.60
SBS0335-052	16.45	15.40	14.02	12.79	1.05	2.38	1.31	7.79	8.72
UM448	14.44	11.33	10.65	10.49	3.11	3.62	0.45	10.30	10.69
UM462	14.42	12.70	12.33	12.20	1.72	2.01	0.26	7.95	8.20

[a] See Apeendix 6.8.2 for a detailed description of the listed quantities.
[b] Assumed $K-[3.6]=0.03$
[c] Assumed $K-[4.5]=0.20$

As previously mentioned, inferring stellar masses photometrically can be problematic for some of the BCDs in our sample. Hot dust, together with free-free nebular continuum or a high equivalent-width Brα line, can contaminate the broadband fluxes from 2 to 5 μm (Hunt et al. 2002, Smith & Hancock 2009). For this reason K, [3.6], and [4.5] magnitudes can potentially be poor indicators of stellar mass. An extreme case is SBS 0335−052 E, one of the lowest metallicity objects in the sample, where 50% of the K band emission is gas, and 13% is dust. Only 37% of the 2 μm emission is stellar (Hunt et al. 2001). At 3.8 μm (ground-based L band), the situation is even worse, with stars comprising only 6% of the emission. Hence, to mitigate the potential overestimate of the stellar mass from contaminated red colors, the minimum (bluest) colors were used to infer the mass-to-light ratio (because of its $B-K$ dependence), and the K-band luminosity. A comparison between the mass-metallicity relation obtained with our inferred stellar masses and the sample in Lee et al. (2006) suggests that the use of the minimum stellar masses (bluest colors) is strongly advocated (with an error of a factor 2-3).

6.8.3 Constraints on the total gas column density

The analysis presented in the main text is entirely based on lower limits on the total gas surface density, because of the impossibility to reliably establish the H_2 abundances from the available CO observations. However, we can impose conservative upper limits on $\Sigma_{\rm gas}$ using indirect ways to quantify $\Sigma_{\rm H_2}$.

For the high-resolution sample, we derive $\Sigma_{\rm gas}$ from the molecular gas column density as inferred by means of the SFRs, assuming a depletion time $t_{\rm depl} \sim 2$ Gyr (Bigiel et al. 2008) for molecular gas. Formally, this should correspond to an upper-limit on $\Sigma_{\rm H_2}$, mainly because we use SFRs integrated on scales which are much greater than the individual associations we are studying. However, since the H I surface density is not precisely known, we cannot regard

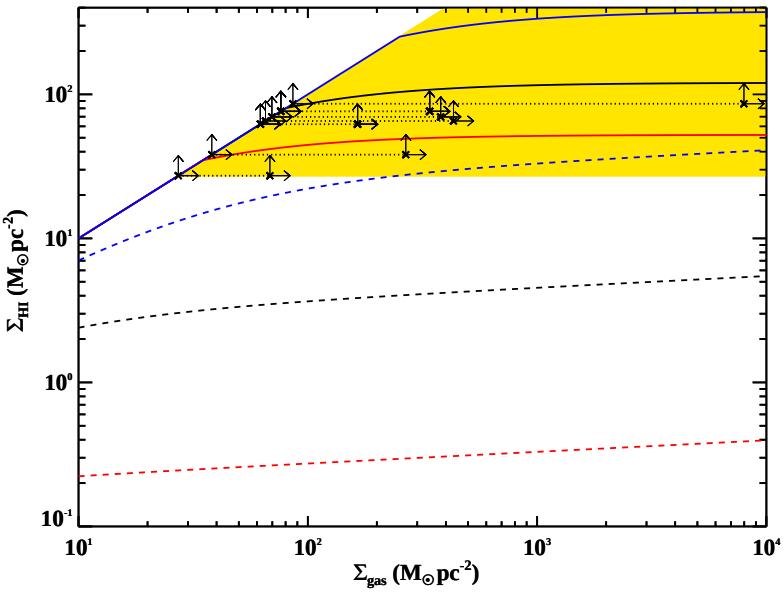

Figure 6.10: Same as Figure 6.4, but with conservative estimates for the total gas column density inferred from SFRs.

$\Sigma_{\rm gas}$ as real upper limits, although we argue that they likely are. Using these limits, we can explicitly show that the disagreement between the extrapolation of the BR model and the observations presented in the Section 6.5.1.3 cannot simply be explained with high gas column densities. This is shown in Figure 6.10, where we present once again both the models and data, adding conservative estimates for $\Sigma_{\rm gas}$, connected with a dotted line. The fact that the derived $\Sigma'_{\rm gas} \sim 2 - 3 \times 10^3$ are not enough to account for the observed discrepancy confirm the results inferred using lower limits only (Figure 6.4).

Similarly, for the low resolution sample, we set conservative upper limits on $\Sigma_{\rm gas}$ assuming a CO-to-H_2 conversion factor $X = 11 \times 10^{21}$ cm^{-2} K^{-1} km^{-1} s (Leroy et al. 2009a). Being derived from one the highest X published to date, the inferred $\Sigma_{\rm H_2}$ are likely to be truly upper limits on the intrinsic H_2. However, whenever these are smaller than $\Sigma_{\rm H_2}$ as obtained from the SFRs combined with a depletion time $t_{\rm depl} \sim 2$ Gyr, we assume conservatively the latter values.

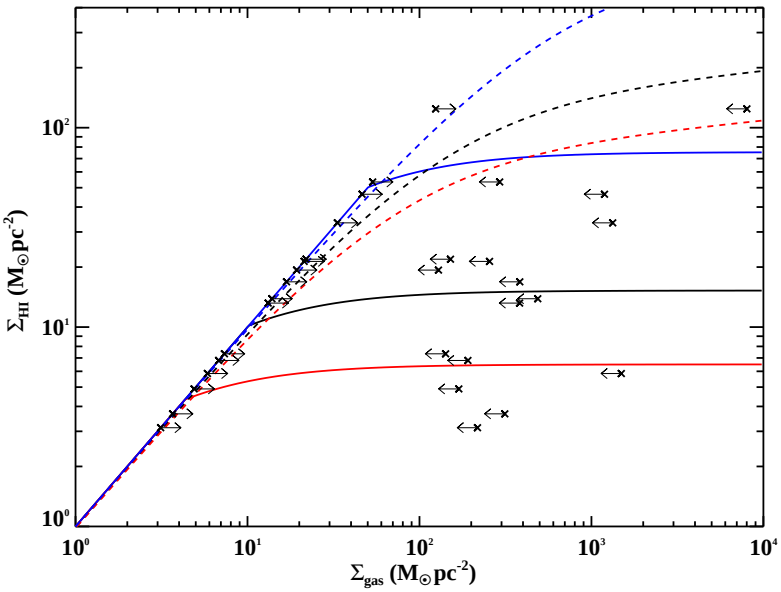

Figure 6.11: Same as Figure 6.5, but with upper limits on the total gas surface density inferred from either CO fluxes or SFRs.

This may be warranted since some galaxies in our sample are at even lower metallicity than the one assumed in CO-to-H_2 conversion factor used here. These upper limits on Σ_{H_2} are shown in Figure 6.11 and can be used in turn to set upper limits on $\Sigma_{HI,mod}$ for the BR model (see Figure 6.6 and Section 6.5.2.2), as in Figure 6.12. Because these upper limits on $\Sigma_{HI,mod}$ exceed significantly the model expectations at low metallicity, we infer that some caution is advisable when extrapolating local empirical star formation laws to high redshift, in dwarf galaxies or in the outskirt of spiral galaxies (see Fumagalli & Gavazzi 2008).

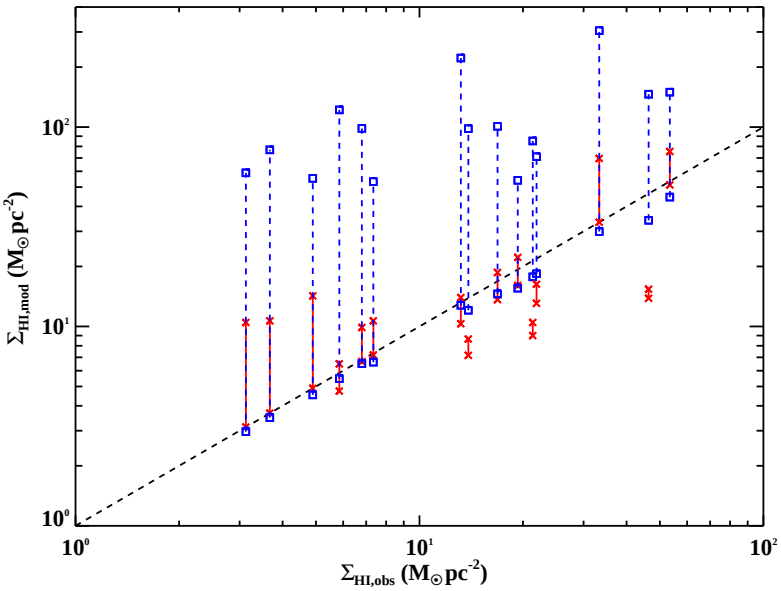

Figure 6.12: Same as Figure 6.6, but with lines to connect lower and upper limits on $\Sigma_{\rm HI,mod}$ for the BR model. Upper limits on $\Sigma_{\rm HI,mod}$ are derived assuming an upper limit on $\Sigma_{\rm gas}$.

Chapter 7

Molecular hydrogen deficiency in HI-poor galaxies and its implications for star formation

This chapter presents a reprint of the paper "Molecular hydrogen deficiency in HI-poor galaxies and its implications for star formation" by Fumagalli et al. (2009) in which we present a second test for the validity of models that describe the conversion of atomic to molecular hydrogen as a function of gas column density and metallicity.

7.1 Abstract

We use a sample of 47 homogeneous and high sensitivity CO images taken from the Nobeyama and BIMA surveys to demonstrate that, contrary to common belief, a significant number ($\sim 40\%$) of HI–deficient nearby spiral galaxies are also depleted in molecular hydrogen. While HI–deficiency by itself is not a sufficient condition for molecular gas depletion, we find

that H_2 reduction is associated with the removal of HI inside the galaxy optical disk. Those HI–deficient galaxies with normal H_2 content have lost HI mainly from outside their optical disks, where the H_2 content is low in all galaxies. This finding is consistent with theoretical models in which the molecular fraction in a galaxy is determined primarily by its gas column density. Our result is supported by indirect evidence that molecular deficient galaxies form stars at a lower rate or have dimmer far infrared fluxes than gas rich galaxies, as expected if the star formation rate is determined by the molecular hydrogen content. Our result is consistent with a scenario in which, when the atomic gas column density is lowered inside the optical disk below the critical value required to form molecular hydrogen and stars, spirals become quiescent and passive evolving systems. We speculate that this process would act on the time–scale set by the gas depletion rate and might be a first step for the transition between the blue and red sequence observed in the color–magnitude diagram.

7.2 Introduction

A complete and coherent description of the detailed connection between a galaxy's gaseous phase and its star formation is still lacking. This poses a significant limitation for studying galaxy evolution, even in the local universe. From an observational point of view, two different approaches have been adopted to address this fundamental topic. The first takes advantage of high spatial resolution images to compare directly the molecular and atomic phases of the interstellar medium (ISM) with recent star formation activities. The results (Wong & Blitz 2002, Kennicutt et al. 2007, Fumagalli & Gavazzi 2008, Bigiel et al. 2008) suggest that stars form preferentially in molecular regions, as traced by diffuse CO emission. The second approach is more indirect, but relies on estimators for the star formation rate and atomic gas that can be easily collected for large samples. These projects statistically study the effects of the

atomic gas removal on star formation in cluster galaxies compared with field galaxies. Results (Kennicutt & Kent 1983, Gavazzi et al. 2002, Koopmann & Kenney 2004, Gallazzi et al. 2009; Gavazzi et. al, in prep.) show a significant correlation between the atomic gas and the star formation: the latter is quenched in those galaxies which are HI deficient due to environmental effects. Although this result can be interpreted in terms of HI feeding star formation (Boselli et al. 2001), it is possible that molecular hydrogen (H_2) depletion also plays a critical role.

However, observations to date suggest that molecular hydrogen, which dominates the mass of the interstellar medium in the center of spiral galaxies, is not strongly affected by environmental perturbations. Galaxies in rich clusters appear to have on average the same molecular content as field galaxies (Stark et al. 1986, Kenney & Young 1986; 1989, Casoli et al. 1991, Boselli et al. 1997; 2002). This result is consistent with a simple theoretical argument based on the assumption that ram-pressure stripping is the dominant environmental process, a fact that, although not yet proved, is likely to be the correct interpretation for HI observations in the Virgo and Coma clusters (see a review by Boselli & Gavazzi 2006). According to this picture, a galaxy which travels at a velocity $v_{\rm rel}$ relative to the dense intergalactic medium (IGM) or intracluster medium (ICM) characterised by a density $\rho_{\rm IGM}$ looses its ISM at the escape velocity $v_{\rm esc}$ and density $\rho_{\rm ISM}$ only if

$$\rho_{\rm IGM} v_{\rm rel}^2 > \rho_{\rm ISM} v_{esc}^2 \, . \tag{7.1}$$

This condition is not easily satisfied for gas in molecular hydrogen phase, which has densities up to 10^5 cm^{-3}, and is usually strongly bound to the galaxy due to its proximity to the galactic center.

This theoretical argument seems to agree with observations of the CO abundance in members of the Virgo and Coma clusters. Using a sample composed of 47 Virgo galaxies, Stark et al. (1986) showed that the molecular fraction CO/HI increases towards the center of

the cluster, identified by the position of M87. This is consistent with the removal of atomic hydrogen only from inner cluster galaxies, without any corresponding change in the molecular hydrogen content. With independent CO measurements, the same conclusion was reached by Kenney & Young (1986; 1989), who studied molecular hydrogen deficiency in datasets of 23 and 41 Virgo galaxies. Their analysis shows that the molecular content has not responded to the HI removal on time scales of 10^9 yr, i.e. it is unperturbed within the Virgo crossing time. Furthermore, Casoli et al. (1991) pointed out that even in the Coma cluster, where the ram–pressure is stronger than in Virgo, there is no correlation between the H_2 mass and the atomic hydrogen content of individual spirals. Surprisingly, using the data collected by Kenney & Young (1989) with updated distances, Rengarajan & Iyengar (1992) found from a comparison of normalized H_2 masses that some deficiency of molecular hydrogen holds among Virgo galaxies, as compared to the field. This result is in contrast with a series of papers by Boselli et al. (1997; 2002), who did not find any correlation between the HI and the H_2–deficiency on a large sample of $\sim$ 250 Virgo and Coma galaxies. Except some extreme cases of galaxies which have their molecular gas depleted and displaced (e.g. Vollmer et al. 2008), there is no strong evidence that cluster member galaxies have statistically less molecular gas than field galaxies.

This seems in contradiction with a coherent picture of the star formation in spirals. If stars really do form primarily in molecular gas, one might expect the quenching in the star formation activity to be connected with a molecular gas deficiency in HI poor galaxies. This hypothesis is partially supported by Fumagalli & Gavazzi (2008), who, using a small sample of 28 images from the Nobeyama CO survey (Kuno et al. 2007), find some hints of a possible depletion in the molecular content (see their Fig. 4 and 6). The aim of this paper is test for H_2–deficiency by using homogeneous and high sensitivity data from the Nobeyama (Kuno et al. 2007), BIMA (Helfer et al. 2003), VIVA (Chung at al., in prep.) and THINGS (Walter et al. 2008) surveys. At first, we compare the integrated molecular and atomic gas mass; then, we

study the distribution of the gas as a function of the galaxy radius and finally we compare both the atomic and hydrogen gas content with indicators for the star formation. While previous studies selected galaxies mainly from the Virgo and Coma clusters, our sample includes also objects from nearby groups. In these diverse environments it is less obvious that ram–pressure is the most efficient process in removing the atomic gas. Therefore, in our analysis we will focus only on the effects produced by the HI removal, without addressing to the details of its origin that is likely connected with a generic environmental perturbations or galaxy feedback.

The paper is organized as follows: the sample and the data reduction are discussed in Sec. 7.3, the analysis is given in Sec. 7.4 and the discussion in Sec. 7.5. In particular, we focus on why previous surveys did not detect the H_2–deficiency (§7.5.1), we explain the process responsible for the molecular gas depletion (§7.5.2) and we discuss the implications of H_2–deficiency on the star formation (§7.5.3). The conclusions are in Sec. 7.6.

7.3 Data Reduction

7.3.1 The sample

Our sample is selected based on the criterion of high sensitivity CO images and homogeneity to minimize the scatter in CO measurements. Because previous CO surveys show a relative flux dispersion which severely limits the H_2 deficiency analysis (see §7.5.1), we limit to the Nobeyama CO survey (Kuno et al. 2007) and the BIMA survey (Helfer et al. 2003) which are shown to agree to within 0.4 dex. 47 massive spiral galaxies in the H–band luminosity range $10^{10} \leq L_{\rm H}/L_{\odot} \leq 10^{11}$ meet our selection criterion, 40 from the Nobeyama and 7 from BIMA. Our sample includes only spiral galaxies since early–type (E/dE) galaxies are known to be extremely gas poor (di Serego Alighieri et al. 2007), while late–type galaxies (Sa–Sc, BCD or Irr/dIrr) are typically gas rich (Gavazzi et al. 2008). Table 7.1 summarizes some observational

properties for our sample. Since we are dealing mainly with the full Nobeyama CO atlas, which has been selected using a far infrared (FIR) emission criterion ($F_{100\mu m} > 10$ Jy), our sample may be biased against CO poor galaxies.

Table 7.1: Properties of the galaxy sample.

Name	R.A. (deg.)	Dec. (deg.)	Dist. (Mpc)	Type	$L_{\rm H}$ ($L_\odot$)	R_{25} (")	$def_{\rm H_2}$	$def_{\rm HI}$	$H\alpha+[NII]$ (E.W. Å)	CO Ref.[a]
IC 342	56.70	68.10	3.9	Sc	10.5	598	-0.24	-0.78	56	N
Maffei 2	40.48	59.60	3.4	Sbc	10.1	174	-0.10	-0.17	-	N
NGC 0253	11.89	-25.29	3.0	Sc	10.6	789	0.10	0.65	16	N
NGC 0628	24.17	15.78	9.7	Sc	10.5	293	0.19	-0.01	35	B*
NGC 1068	40.67	-0.01	14.4	Sb	11.1	184	-0.11	0.47	50	N
NGC 2903	143.04	21.50	6.3	Sb	10.4	352	0.02	0.33	18	N*
NGC 3184	154.57	41.42	8.7	Sc	10.2	222	0.11	0.32	-	N*
NGC 3351	160.99	11.70	8.1	Sb	10.4	217	0.34	0.60	16	N*
NGC 3504	165.80	27.97	26.5	Sab	10.7	73	0.10	0.37	35	N
NGC 3521	166.45	-0.03	7.2	Sb	10.6	249	0.19	0.06	20	N*
NGC 3627	170.06	12.99	6.6	Sb	10.5	306	-0.11	0.81	19	N*
NGC 3631	170.26	53.17	21.6	Sc	10.7	111	0.01	-0.03	34	N
NGC 3938	178.20	44.12	17.0	Sc	10.5	106	-0.06	-0.24	26	B
NGC 4051	180.79	44.53	17.0	Sb	10.6	146	-0.03	0.26	32	N
NGC 4102	181.60	52.71	17.0	Sb	10.5	88	-0.02	0.47	26	N
NGC 4192	183.45	14.90	17.0	Sb	11.0	293	0.59	0.36	9	N
NGC 4212	183.91	13.90	16.8	Sc	10.3	108	0.17	0.57	20	N
NGC 4254	184.71	14.42	17.0	Sc	10.9	184	-0.20	0.18	32	N
NGC 4258	184.74	47.30	6.8	Sb	10.7	545	0.44	0.25	15	B
NGC 4303	185.48	4.47	17.0	Sc	11.0	197	-0.08	0.13	36	N
NGC 4321	185.73	15.82	17.0	Sc	11.1	273	0.05	0.53	18	N*
NGC 4402	186.53	13.11	17.0	Sc	10.4	118	0.02	0.73	12	N
NGC 4414	186.61	31.22	17.0	Sc	10.9	108	-0.09	-0.15	23	N
NGC 4419	186.74	15.05	17.0	Sa	10.6	105	0.40	1.04	7	N*
NGC 4501	188.00	14.42	17.0	Sbc	11.2	216	0.14	0.51	6	N
NGC 4535	188.58	8.20	17.0	Sc	10.8	249	0.02	0.35	17	N
NGC 4536	188.61	2.19	17.0	Sc	10.7	216	0.18	0.29	20	N*
NGC 4548	188.86	14.50	17.0	Sb	10.9	180	0.84	0.70	17	N
NGC 4569	189.21	13.16	17.0	Sab	11.0	321	0.23	1.13	2	N*

Continued on Next Page...

Table 7.1 – Continued

Name	R.A. (deg.)	Dec. (deg.)	Dist. (Mpc)	Type	$L_{\rm H}$ ($L_\odot$)	R_{25} (")	$def_{\rm H_2}$	$def_{\rm HI}$	$H\alpha + [NII]$ (E.W. Å)	CO Ref.[a]
NGC 4579	189.43	11.82	17.0	Sab	11.1	188	0.62	0.71	4	N
NGC 4654	190.99	13.13	17.0	Sc	10.7	149	-0.02	0.17	31	N*
NGC 4689	191.94	13.76	17.0	Sc	10.5	175	0.14	1.06	14	N*
NGC 4736	192.72	41.12	4.3	Sab	10.4	232	0.78	0.68	10	N*
NGC 4826	194.18	21.68	4.1	Sab	10.3	314	0.70	0.86	9	B*
NGC 5005	197.73	37.06	21.3	Sb	11.3	143	0.28	0.47	7	B
NGC 5033	198.36	36.59	18.7	Sc	11.0	293	-0.16	0.19	20	B
NGC 5055	198.96	42.03	7.2	Sbc	10.7	352	0.04	0.12	20	N*
NGC 5194	202.47	47.19	7.7	Sbc	10.8	232	-0.36	0.12	28	N*
NGC 5236	204.25	-29.86	4.7	Sc	10.7	464	0.23	-0.49	33	N*
NGC 5247	204.51	-17.88	22.2	Sb	10.9	161	0.06	0.05	-	N
NGC 5248	204.38	8.88	22.7	Sb	11.1	122	-0.10	-0.25	26	N
NGC 5457	210.80	54.35	5.4	Sc	10.5	719	0.06	0.28	31	N*
NGC 6217	248.16	78.20	23.9	Sbc	10.5	67	0.13	-0.45	34	N
NGC 6946	308.72	60.15	5.5	Sc	10.4	344	-0.28	-0.45	34	N*
NGC 6951	309.31	66.10	20.3	Sb	11.1	97	0.24	0.04	16	N
NGC 7331	339.27	34.42	14.3	Sbc	11.1	273	0.14	-0.04	-	B*
UGC 2855	57.09	70.13	20.3	Sc	10.9	132	-0.34	-0.13	34	N

[a] N: Nobeyama survey (Kuno et al. 2007); B: BIMA survey (Helfer et al. 2003). Galaxies included in the high–resolution study are identified by an asterisk.

The Nobeyama maps are obtained with a 45m single–dish radio telescope that provides high sensitivity ($\sim$ 0.5 $M_\odot$ pc^2) flux measurements for the entire galaxy disk. The spatial resolution is about $15''$, which corresponds to an average physical scale of $\sim$ 1 kpc at the mean distance of our sample ($\sim$ 15 Mpc). In contrast, despite its higher spatial resolution ($6'' - 10''$, corresponding to 450–700 pc), the interferometric BIMA survey has systematically poorer sensitivity to extended flux because interferometers cannot detect flux from regions larger than the size–scale set by the minimum distance between the antennas. One can recover the total flux by using single dish observations that provide the zero spacing element. We have limited our galaxy sample to the subset with single–dish data. Even for these galaxies, the reconstruction algorithm is biased against large scales (Helfer et al. 2003; app. A); however, a comparison between Nobeyama and BIMA data indicates that the missed flux is negligible compared to the total flux, i.e. the BIMA survey provides a reliable estimate of the integrated molecular hydrogen mass. When a galaxy is imaged in both the surveys, we use the Nobeyama data.

For our analysis we will also compare the spatial distributions of molecular and atomic hydrogen; therefore we collect spatially resolved HI images from the THINGS survey (Walter et al. 2008) and the VIVA survey (Chung et al., in prep.). Since spatially resolved images are not available for all the galaxies in our sample, while studying surface density profiles we restrict our analysis to a subsample of 20 galaxies (see Table 7.1).

7.3.2 A multifrequency analysis

The comparison between the molecular distribution, the atomic hydrogen content and the star formation activity requires some preliminary transformations from observed to physical quantities. In this section we describe this issue.

7.3.2.1 The molecular hydrogen

The H_2 molecule has no permanent dipole so we must infer its abundance indirectly by assuming that the conditions of the ISM where the H_2 forms are similar to those in which other molecules are synthesized. A good tracer for H_2 is CO because the intensities of its rotational transitions are correlated to the molecular column density. This is based on a simple argument which relates the velocity inferred from the line profile with the virial velocity of molecular clouds (Dickman et al. 1986, Solomon et al. 1987).

The Nobeyama and BIMA surveys image the CO $J : 1 - 0$ transition at 115 GHz, which is excited above 5.5 K in cool gas at moderate density. We infer the molecular hydrogen column density by adopting a conversion factor X (cm^{-2} $(\mathrm{K\ km/s})^{-1}$), which is a function of temperature, density, UV radiation field, metallicity and even of the shape of the molecular clouds (Wall 2007, Boselli et al. 2002; 1997; app. A). Because it is impossible to model X theoretically for actual galaxies, a constant conversion factor derived using observations in our Galaxy has often been assumed (e.g. Solomon et al. 1987). The Galactic factor likely gives underestimates for the molecular hydrogen content in low mass galaxies, because empirical calibrations show that X decreases for increasing mass (e.g. Boselli et al. 2002). Since we expect the CO–to–H_2 conversion factor to depend primarily on the dust and metallicity that regulate the formation and photodissociation of H_2 and the CO heating, we choose an individual X value for each galaxy following the empirical calibrations of Boselli et al. (2002). This is

$$\log X = -0.38 \log L_{\mathrm{H}} + 24.23\,, \tag{7.2}$$

where L_{H} is the H–band luminosity in $\mathrm{L}_{\odot}$. We retrieve L_{H} from GOLDMine (Gavazzi et al. 2003) and SIMBAD. This calibration accounts for the dependence of X on the metal abundance because the H–band luminosity correlates with galaxy metallicity (e.g. Boselli et al. 2002). The

systematic uncertainty on the X factor derived using eq. (7.2) ranges from 0.65 to 0.70 dex for the luminosity interval $10^{10} - 10^{11}$ $L_{\odot}$. Other unknown dependencies can increase this error. The final conversion for the molecular hydrogen surface density is:

$$\Sigma_{H_2} \, [\mathrm{M}_{\odot} \, \mathrm{cm}^{-2}] = 2 m_p X F_{\mathrm{co}} \, , \tag{7.3}$$

where m_p is the proton mass in $M_{\odot}$ and F_{co} is the CO flux in K km/s. This conversion does not include the mass of helium, but its contribution may be included trivially by multiplying the derived surface density by 1.36. The integrated CO flux ($\bar{F}_{\mathrm{co}}$ in Jy km/s) is converted into a total H_2 mass ($M_{\mathrm{H_2}}$ in $M_{\odot}$) using (Helfer et al. 2003):

$$M_{H_2} = 3.92 \times 10^{-17} X \bar{F}_{\mathrm{co}} D^2 \, , \tag{7.4}$$

where D is the galaxy distance in Mpc from GOLDMine and Tully (1994).

7.3.2.2 The atomic hydrogen

We estimate the atomic hydrogen content via 21 cm emission line. Single–dish HI surveys, including the ALFALFA (Giovanelli et al. 2005) or HIPASS (Meyer et al. 2004) surveys, do provide the HI integrated flux for all the galaxies in our sample. Collecting different HI data for a same galaxy, we notice a significant dispersion in the measured HI flux, even up to a factor of 2. Because we are interested in a relative comparison between galaxies rather than an accurate determination of the absolute HI mass, we retrieve all the HI fluxes from the HyperLeda database (Paturel et al. 2003a). Even if more recent measurements have better sensitivity and provide a more accurate determination of the total HI abundance, this approach gives an homogeneous dataset (Paturel et al. 2003b). As discussed in Appendix 7.7.1, a particular choice for the

observed flux does not affect our analysis.

The integrated HI flux ($\bar{F}_{\rm HI}$ in Jy km/s) is converted into a total mass ($M_{\rm HI}$ in $\rm M_\odot$) using

$$M_{\rm HI} = 2.36 \times 10^5 \bar{F}_{\rm HI} D^2 \, , \tag{7.5}$$

where D is the galaxy distance in Mpc from GOLDMine and Tully (1994). The HI surface density profile is computed from THINGS and VIVA images according to the conversion

$$\Sigma_{HI} \, [\rm M_\odot \, cm^{-2}] = m_p N_{\rm HI} \, , \tag{7.6}$$

where $N_{\rm HI}$ is the column density in $\rm cm^{-2}$ derived from the flux $F_{\rm HI}$ (Jy/beam km/s) using the following calibration (e.g. Walter et al. 2008):

$$N_{\rm HI}[10^{19}{\rm cm}^2] = 110.4 \times 10^3 \frac{F_{\rm HI}}{B_{\rm min} \times B_{\rm max}} \tag{7.7}$$

with B the beam FWHM in arcsec. As with $\rm H_2$, these surface densities do not include the mass of helium.

7.3.2.3 The star formation activity

In the ISM of galaxies, the $H\alpha$ emission line is caused by the recombination of atomic hydrogen which has been ionized by the UV radiation of O and B stars. Therefore $H\alpha$ is a tracer of recent, massive ($M > 10$ $\rm M_\odot$) star formation activity on time scales of $t \leq 4 \times 10^6$ yr (Kennicutt 1998). For all but three objects, we retrieve from the literature (GOLDmine; Kennicutt & Kent 1983, James et al. 2004, Kennicutt et al. 2008) the $H\alpha$ equivalent width (E.W.), a tracer of the specific star formation rate (Kennicutt 1983b).

Since the narrow band filter used to image the $H\alpha$ emission line is large enough to

collect the flux from the $[NII]$ line, the measured E.W. overestimates the $H\alpha$ emission. Unfortunately, the ratio $[NII]/H\alpha$ is not available for all of the galaxies in our sample and we can apply only the average correction $[NII]/H\alpha = 0.53$, as suggested by Kennicutt (1992). However, this value underestimates the $[NII]$ emission from the bulges of galaxies that host active galactic nuclei (AGN), where the nuclear $[NII]$ can be even stronger than the $H\alpha$. In addition, even if the nuclear emission were corrected using a suitable factor, it would not be possible to disentangle between the $H\alpha$ flux due to star formation from that due to AGN ionization. Our sample, composed of massive spirals, is likely to be affected by this contamination (Decarli et al. 2007).

Another effect to consider is dust absorption. In principle one can estimate the extinction by measuring the $H\alpha/H\beta$ or the $H\alpha/Br_\gamma$ fluxes (case B recombination). These values, however, are unknown for the majority of our galaxies. One could apply a standard dust extinction (e.g. $A = 1.1$ mag, Kennicutt & Kent 1983) but this would not change the relative distribution of the measurements. Instead, we simply keep the measured $H\alpha + [NII]$ E.W. as an approximation for the specific star formation rate. The reader should keep in mind that a large uncertainty is associated with these measurements, in addition to the error in the flux determination.

7.3.3 Global fluxes and surface brightness profiles

We compute total CO fluxes from the Nobeyama images using the IRAF task `qphot` by integrating the emission in an aperture which encloses the entire galaxy. Fluxes from the BIMA survey are retrieved from Helfer et al. (2003), who list values computed within a square region. We retrieve the HI flux from HyperLeda, as previously discussed. We compute the surface brightness profiles for the CO and HI using an IDL procedure that averages the flux into elliptical annuli characterised by three fixed parameters (the galaxy center, the eccentricity

and the position angle); we adopt a step along the major axis of ~ 5 arcsec, roughly 1/3 of the spatial resolution of the CO images from the Nobeyama survey. All the profiles are corrected for the projection effect using the galaxy inclination (HyperLeda) on the plane of the sky. For each galaxy we finally obtain the HI and H_2 surface density profiles Σ_{HI} and Σ_{H_2} in $M_\odot$ pc^{-2}.

7.4 Analysis

To test for molecular hydrogen depletion in our dataset, we compare the H_2 mass in a sample of isolated galaxies with that in perturbed galaxies. Our isolation criterion is defined by the atomic hydrogen content. Because the edges of the HI disk are weakly bounded, any perturbation can easily remove atomic gas from a galaxy, lowering its HI mass. That the HI–deficiency should trace environmental effects is supported by Giovanelli & Haynes (1985), who show that HI–deficiency increases toward the center of a cluster. In these regions the IGM is dense, the cluster potential well is deep and the number of galaxies is high so that environmental processes are very effective. We expect HI deficiency to be a useful measure of disturbance even if environment is not the dominant means of gas removal. For example, if star formation feedback is the primary cause of gas loss, this too will preferentially remove the low–density, loosely bound HI.

Formally, the atomic gas depletion is quantified using $def_{\rm HI}$, the deficiency parameter, defined by Haynes & Giovanelli (1984) as the logarithmic difference between the expected HI mass in isolated galaxies and the observed value:

$$def_{\rm HI} = \log M_{\rm HI,exp} - \log M_{\rm HI,obs} \tag{7.8}$$

The reference mass $M_{\rm HI,exp}$ is computed as a function of the optical radius at the 25th mag/arcsec^2

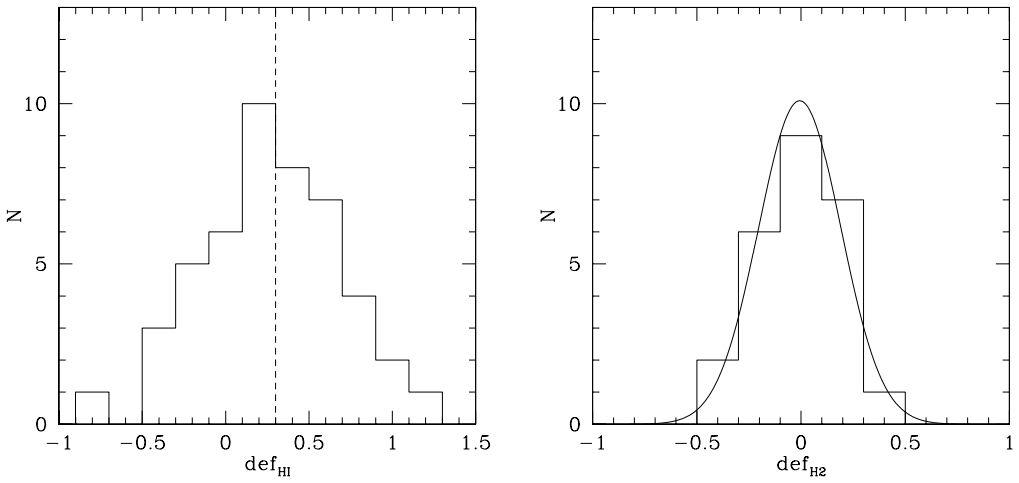

Figure 7.1: a - Histogram of the HI–deficiency values in our sample, separated in isolated and perturbed galaxies at $def_{\rm HI} = 0.3$. A positive value of $def_{\rm HI}$ indicates that a galaxy has an HI mass lower than expected on average according to its size and morphology. b - Histogram of the H_2–deficiency values for isolated galaxies only. The Gaussian fit shows that $\sigma = 0.2$.

in B band (retrieved from NED and HyperLeda) using the coefficients recently updated by Solanes et al. (1996), which are weakly dependent on the morphological type. Several sources of uncertainty affect the determination of $def_{\rm HI}$: the error on the observed HI mass, the statistical uncertainty on the calibrated coefficients and the determination of the optical diameter. It is hard to quantify a value for the final uncertainty, but a comparison of the values derived for a single galaxy using different data available in the literature suggests a mean dispersion of $0.2 - 0.3$ dex on the HI–deficiency.

Figure 7.1a shows the distribution of the HI–deficiencies for our sample. If we arbitrarily consider unperturbed those galaxies which have lost less than a factor of 2 in their HI mass ($def_{\rm HI} < 0.3$), our subsample of isolated galaxies is composed of 25 objects with a mean deficiency $def_{\rm HI} = -0.045 \pm 0.27$. This is consistent with zero. Moreover, we point out that the dispersion on the mean value is within the deficiency threshold of 0.3 dex. Only 4 galaxies lie at $def_{\rm HI} < -0.3$, indicating that our reference sample is not dominated by galaxies with HI mass

above the average. The remaining 22 galaxies have $def_{\rm HI} \geq 0.3$, spanning the entire range of moderate and high deficiencies ($0.3 - 1.2$ dex).

In a similar way, we quantify a possible reduction in the molecular hydrogen content using the H_2–deficiency parameter $def_{\rm H_2}$ (Boselli et al. 1997; 2002):

$$def_{\rm H_2} = \log M_{\rm H_2,exp} - \log M_{\rm H_2,obs} \, . \tag{7.9}$$

The value for the reference mass $M_{\rm H_2,exp}$ is computed from the observed correlation between the H_2 and the H–band luminosity (Boselli et al. 1997) for isolated galaxies ($def_{\rm HI} < 0.3$). This definition is based on the natural scaling of the molecular gas abundance with the size of a galaxy, i.e. more massive spiral galaxies are more gas rich. Although our sample covers a small range in luminosity, owing to the good quality of our data (see section 7.5.1), we choose to compute a new reference mass with a least square fit:

$$\log M_{\rm H_2,exp} = 0.50 \log L_{\rm H} + 3.98 \, . \tag{7.10}$$

Since the dispersion (rms) in the calibration is ~ 0.2 (See also Fig. 7.1b), galaxies with $def_{\rm H_2} < 0.2$ cannot be distinguished from normal galaxies. In the following analysis we therefore consider a galaxy H_2–deficient only if $def_{\rm H_2} \geq 0.2$. However, as we shall see below, $def_{\rm H_2}$ shows a continuous rather than a bimodal distribution. Thus, this criterion is arbitrary. It does have the virtue, though, of nicely dividing our sample into classes which have a comparable number of objects. As with $def_{\rm HI}$, it is hard to quantify an uncertainty for $def_{\rm H_2}$. Deficiency is a relative quantity, so the level of uncertainty for $def_{\rm H_2}$ is likely to be significantly lower than for the absolute H_2 mass. The error of 0.7 dex we quoted in equation (7.2) is therefore an upper limit on the scatter in $def_{\rm H_2}$. A more reliable estimate of the error is the dispersion of 0.2 dex we measure

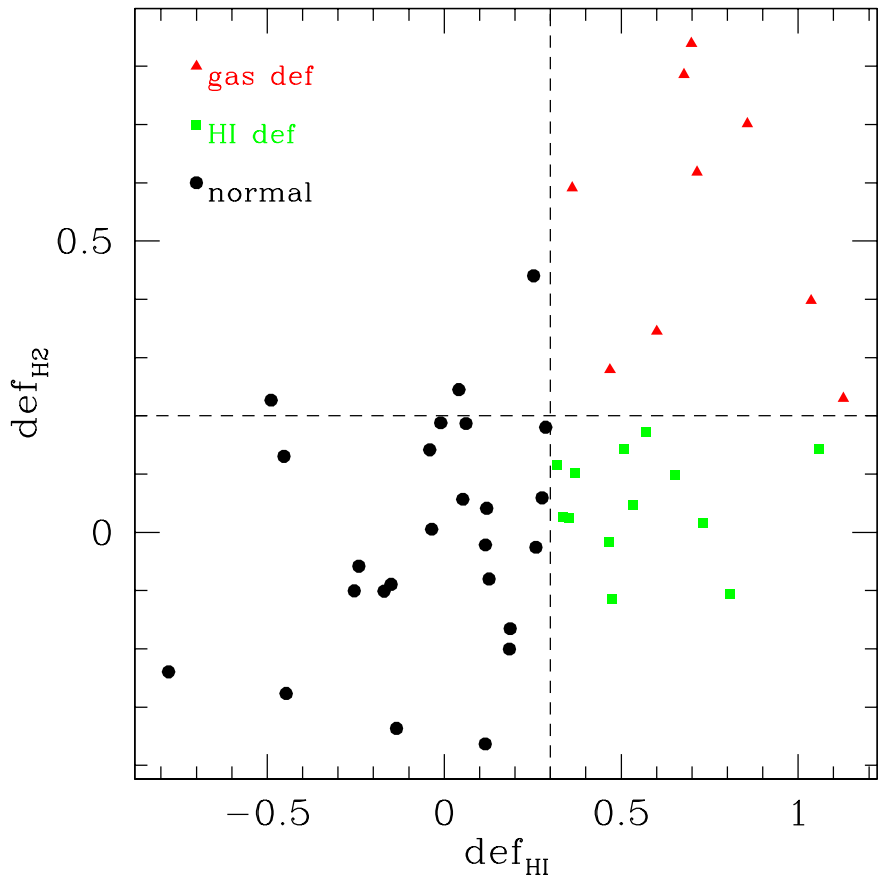

Figure 7.2: Comparison between calculated $def_{\rm H_2}$ and $def_{\rm HI}$ values for our sample. The vertical dashed line corresponds to $def_{\rm HI} = 0.3$ dex, the value below which a galaxy is considered unperturbed; the horizontal line represents $def_{H_2} = 0.2$ dex. The absence of galaxies in the upper–left quadrant suggests that HI–deficiency is only a necessary condition for H_2–deficiency.

in $def_{\rm H_2}$ for the control sample of isolated galaxies (Fig. 7.1b). Since random error in computing M_{H_2}, for example galaxy–to–galaxy scatter in X, can only broaden the intrinsic dispersion of H_2 masses at fixed L_H, our measured scatter of 0.2 dex sets a more realistic upper limit on the size of random errors. The one caveat is that this value may neglect systematic uncertainties if unperturbed and perturbed galaxies have different values of luminosity or metallicity, for example, and thus systematically different values of X. However, we do not observe or expect any obvious trend of these quantities as a function of $def_{\rm HI}$ or $def_{\rm H_2}$.

Figure 7.2 shows a comparison between $def_{\rm H_2}$ and $def_{\rm HI}$ in our sample. The vertical dashed line corresponds to $def_{\rm HI} = 0.3$ dex, the value below which a galaxy is considered unperturbed; the horizontal line represents $def_{\rm H_2} = 0.2$ dex. Inspecting this plot, the first striking result is that 9/22 of the galaxies with $def_{\rm HI} \geq 0.3$ have $def_{\rm H_2}$ values greater than 0.2 dex, while the region characterised by $def_{\rm HI} < 0.3$ and $def_{\rm H_2} \geq 0.2$ is almost empty. This distribution proves the existence of the molecular deficiency in a significant number of galaxies which are HI poor. This figure indicates that $def_{\rm H_2}$ and $def_{\rm HI}$ do not follow a tight correlation, but this is not surprising. In fact, we do not expect effects that deplete the HI to act in the same way on the H_2 gas which is more strongly bound. Rather than a correlation, it is evident that galaxies lie in three regions roughly defined by our criterion of isolation and molecular deficiency (dashed lines). Based on Fig. 7.2, we divide our sample in three classes: *unperturbed* or *non–deficient* galaxies ($def_{\rm HI} < 0.3$ and $def_{\rm H_2} < 0.2$), *gas–deficient* galaxies ($def_{\rm HI} \geq 0.3$ and $def_{\rm H_2} \geq 0.2$) and galaxies which are only *HI–deficient* ($def_{\rm HI} \geq 0.3$ and $def_{\rm H_2} < 0.2$). This classification should not be considered strict, since galaxies are distributed continuously rather than in distinct populations, but it is useful for the analysis. We test the robustness of our result by changing the isolation criterion in the range $def_{\rm HI} = 0.1 - 0.4$ and recomputing $def_{\rm H_2}$ after that a new calibration has been derived. We find that the distribution in Figure 7.2 does not change significantly.

Since the definition of the molecular gas deficiency relies on Eq. (7.10) which sets the reference mass, the behaviour shown in Fig. 7.2 could be an artifact imposed by this calibration. We further explore the robustness of a H_2–deficient population by studying the H_2–deficiency with surface density gas profiles, which is independent from the calibration of $\log M_{\rm H_2,exp}$. In Figure 7.3 we show the average molecular gas surface density profiles, in bins of normalized galactocentric radius[1] (R/R_{25}). Not all the galaxies in our sample are imaged in HI. Therefore,

[1]This quantity is well defined under the hypothesis that the stellar disk is not truncated during the pertur-

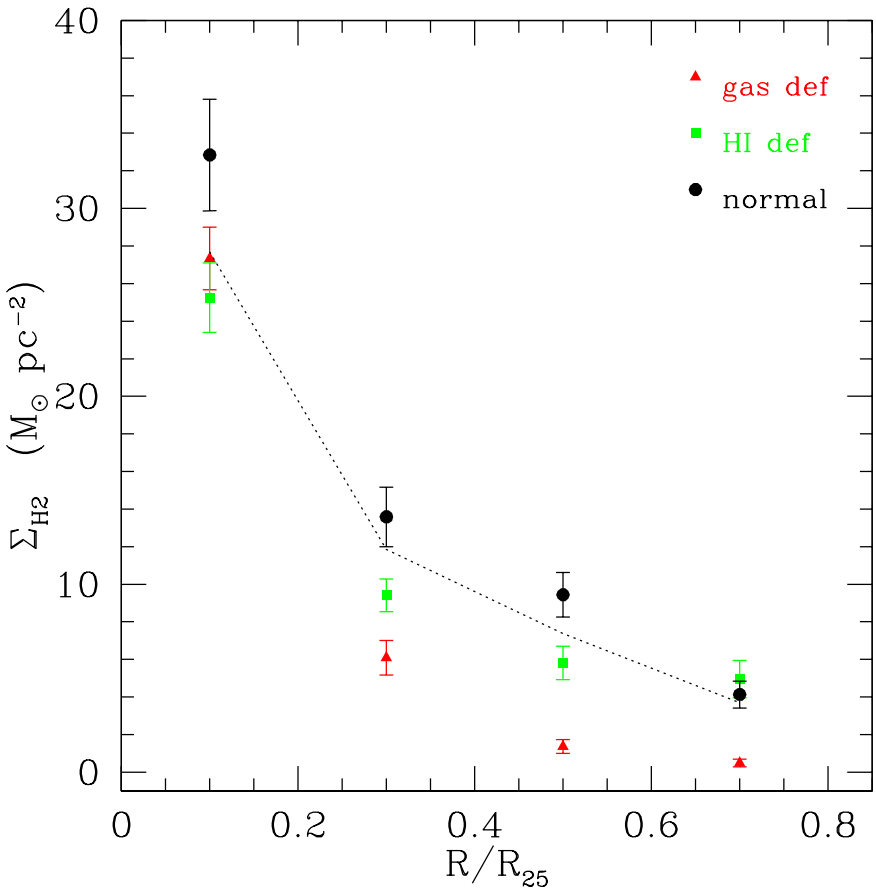

Figure 7.3: Average H_2 profiles in bins of normalized radius. The dashed line is computed by excluding HI–rich galaxies with $def_{\rm HI} < -0.3$. The offset among the profiles reflects the def_{H_2} distribution in Fig. 7.2.

for this kind of analysis, we consider only 20 galaxies (See Table 7.1). Details on the completeness in each class are provided in Table 7.2. By comparing the values for the three classes of galaxies (non–deficient with circles, gas–deficient with triangles and HI–deficient with squares), it is evident that gas–deficient galaxies have the lowest molecular gas column densities, while HI–deficient and non–deficient galaxies exhibit a similar behaviour. The dashed line in Fig. 7.3 shows the average profile excluding those galaxies with $def_{\rm HI} < -0.3$. The similarity of this profile to that of the full non–deficient sample (circles) indicates that the observed difference in H_2 content between the three galaxy classes is not an artifact enhanced by the presence of gas–rich galaxies in our sample. Based on an independent method, Fig. 7.3 confirms the existence of H_2–deficient galaxies in $\sim 40\%$ of the HI poor galaxies and $\sim 20\%$ of all the galaxies in our sample. As discussed in the next section, these values should be regarded as lower limits.

Table 7.2: Completness for the surface density profiles.

Class	Objects	Class[a]	Total[b]	Symbol[c]
Non–deficient	10	40%	21%	Circle
Non gas–rich	8	38%	17%	Line
HI–deficient	5	38%	11%	Square
Gas–deficient	5	56%	11%	Triangle

bation. Clearly this is not the case for tidal interactions. In our sample, NGC 3351 is the galaxy most likely to be affected by tidal stripping. However, we note that a value for R_{25} larger than what we have adopted would shift the surface density profile at lower R/R_{25}, enhancing the separation between unperturbed and gas–deficient galaxies.

[a] % Class is defined as the number of galaxies with HI maps in each class over the total number of objects in that class.

[b] % Total is defined as the number of galaxies with HI maps in each class over 47, the number of objects in the entire sample.

[c] Symbols used for each class in the plots throughout the paper.

7.5 Discussion

7.5.1 Comparison with previous studies

The first issue that we would like to address is why previous studies, based on large CO surveys, did not reveal a population of H_2-deficient galaxies. As discussed in the introduction, several studies (Stark et al. 1986, Kenney & Young 1986; 1989, Casoli et al. 1991, Boselli et al. 1997; 2002) based on large samples both in the Virgo and in the Coma clusters did not show any evidence for a molecular gas deficiency. As discussed by Boselli & Gavazzi (2006), some biases influence the previous works; for example, a constant X conversion factor was often assumed or samples were selected according to FIR selection criteria which prefer CO rich galaxies. Of course, our sample has also been selected to have IRAS fluxes from the Point Source Catalogue higher than 10 Jy at 100 μm. By comparing the 100μm flux with our $def_{\rm H_2}$ in Figure 7.4, we notice that gas–deficient galaxies show systematically lower 100μm luminosity; therefore, we might miss a significant fraction of molecular poor galaxies.

Based on this work, we can identify two additional reasons for why previous studies failed to detect the molecular deficiency: the completeness in the flux sampling and the homogeneity of the data. While observing at a single–dish, the integration time required to detect the CO flux is very high (120 minutes at a 12m telescope are needed to reach a sensitivity of 3 mK at a resolution of 15 km/s) and several observations are required to image the entire galaxies. It has only recently become possible produce CO images in a reasonable observing time, thanks to the the availability of multibeam receivers that reduce the number of pointings required to fully map the entire disk. Due to these technical limitations, previous studies were based on the observation of one position per galaxy plus one or two positions off-center, along the major axis. The total CO flux was then recovered by interpolating with an exponential profile, a procedure that inevitably gives only a rough estimate of the proper H_2 mass. By comparing the integrated

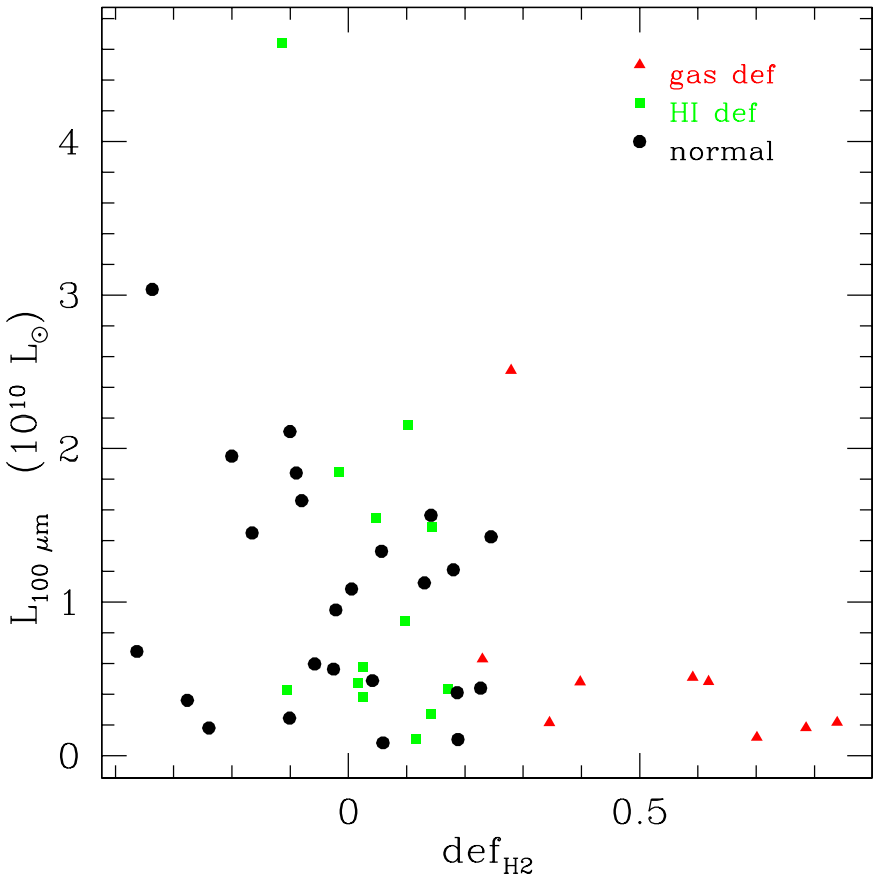

Figure 7.4: Far infrared luminosity at 100μm as a function of the molecular gas deficiency. Gas–deficient galaxies show the lowest 100μm luminosity.

CO flux from Kenney & Young (1988) and GOLDMine with the Nobeyama and BIMA data, we find a dispersion of up to 1 dex. In contrast, by comparing galaxies in common in the BIMA and Nobayama surveys, the observed discrepancy in the total mass is less than 0.4 dex. Therefore, we believe that the data used in previous studies suffer from fluctuations larger than the observable H_2–deficiency (< 0.8 dex in our sample). Another possible problem is related to the homogeneity of the sample. To increase the statistics, earlier studies often combined CO data from surveys with different sensitivities and diverse techniques to recover the integrated flux, increasing the relative dispersion in the observed H_2 mass. We attribute the fact that the H_2–deficiency has remained hidden in previous analysis to the combination of these two effects. On the contrary, based on high quality and homogeneous data from the BIMA and Nobeyama surveys, our work allows for a determination of molecular gas deficiency.

7.5.2 The observed distribution in the H_2–deficiency

The observed distribution in $def_{\rm H_2}$ deserves a detailed analysis. In particular, we are interested in understanding why HI–deficiency is a necessary but insufficient condition for H_2–deficiency and what the fundamental difference is between galaxies which are only HI–deficient and galaxies which are also molecular gas deficient. The answers to these questions lie in the physics that governs the H_2 formation. Molecular hydrogen is synthesized in spiral disks through the interaction of atomic hydrogen with dust grains, whereas it is destroyed by the UV radiation field. The balance between these processes determines the ratio of molecular to atomic hydrogen. In a series of recent papers, Krumholz et al. (2008; 2009a) investigate the problem of atomic to molecular ratio with a model that includes the treatment of H_2 formation, its self-shielding and the dust shielding. According to their prescription, the molecular fraction in a galaxy is determined primarily by the product of its hydrogen column density and metallicity. Under the assumption of a two–phase medium (Wolfire et al. 1995), the results are largely independent

on the UV flux because a change in the UV flux produces a compensating change in the HI density. Hence, the effects of UV radiation field are largely cancelled out. For the purposes of our analysis, the most important prediction of the model by Krumholz et al. is that the fraction of the gas in molecular form increases as the total gas column density $\Sigma_{gas} = \Sigma_{\rm HI} + \Sigma_{\rm H_2}$ and the metallicity Z increase, providing a large enough shield against the photodissociation. In agreement with what is observed in local galaxies, the model predicts that the HI column density is expected to saturate at a value of ~ 10 $M_\odot$ pc^{-2}. Based on the works by Krumholz et al., it is natural to study the behaviour of the total gas column density in our sample to understand the nature of the H_2–deficiency.

Figure 7.5 shows the average total gas surface density profiles $\Sigma_{\rm HI} + \Sigma_{\rm H_2}$ in bins of the normalized galactocentric radius. Inside the optical radius, the total gas content of the HI–deficient galaxies (squares) is the same as that of the unperturbed galaxies (circles), while gas–deficient galaxies (triangles) have a lower total gas column density. This behaviour reflects the observed atomic and molecular gas content seen in Fig. 7.2, yet add new insights into the nature of the H_2 deficiency. Gas–deficient galaxies have a gas column density too low to shield molecules from photodissociation, while HI–deficient galaxies have normal gas content inside their optical radii, which allows the production of molecular hydrogen as in unperturbed galaxies. Indeed, gas–deficient galaxies lie below the critical shielding column density of ~ 10 $M_\odot$ pc^2 while unperturbed and HI–deficient galaxies lie above this threshold. As in Fig. 7.3, we also plot the average gas profile excluding galaxies with $def_{\rm HI} < -0.3$ (dashed line): the consistency of the profiles with or without HI–rich galaxies excludes any bias in the mean value towards high column density. Inspecting the individual HI profiles, we noticed that NGC 4321 exhibits an unusual high column density. Excluding this object from the analysis (open squares) does not change the results beacause the major contribution to the total gas column density comes from the molecular hydrogen.

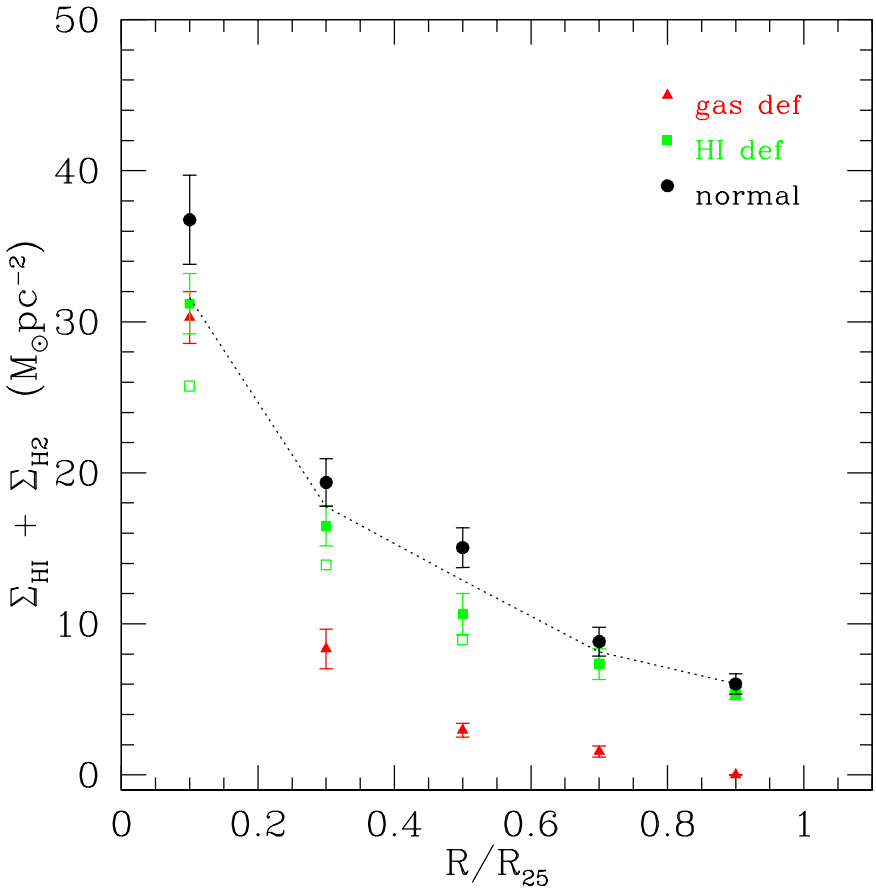

Figure 7.5: Average total gas surface density profiles in bins of normalized radius, divided according to their $def_{\rm HI}$ and def_{H_2} values. The dashed line is computed excluding HI rich galaxies with $def_{\rm HI} < -0.3$ and the open squares indicate the profile for HI–deficient galaxies without NGC 4321. Gas–deficient galaxies have in the disk a gas column density below the values required to form molecular hydrogen (~ 10 M$_\odot$ pc^{-2}).

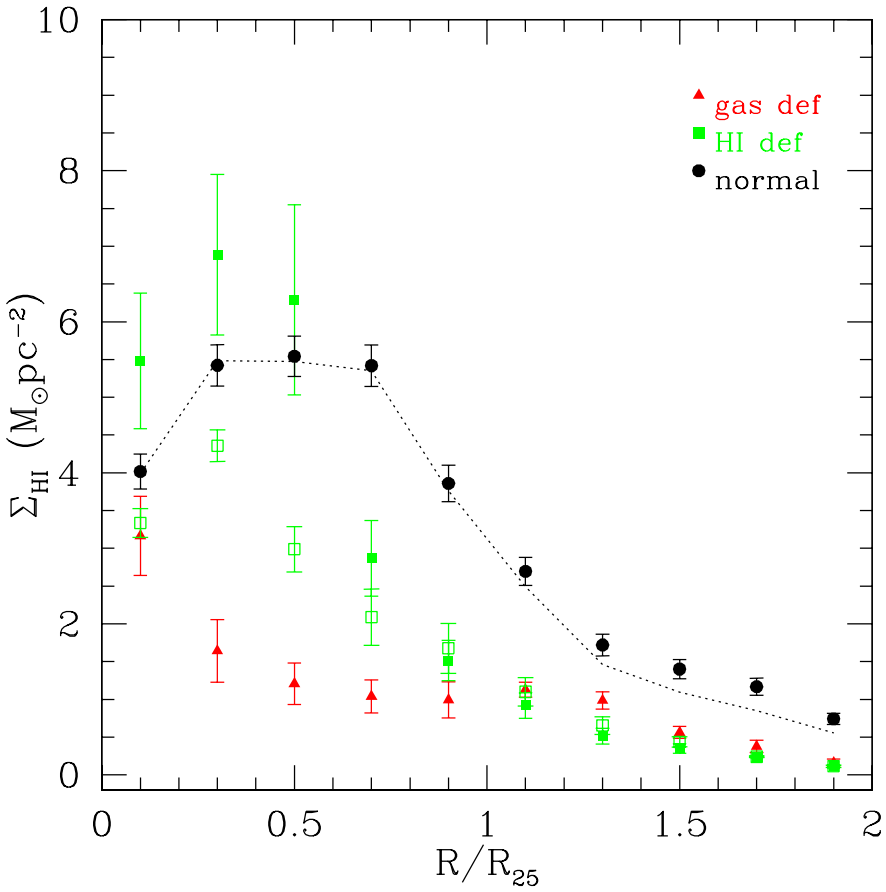

Figure 7.6: Average HI surface density profiles in bins of normalized radius, divided according to their $def_{\rm HI}$ and def_{H_2} values. The dashed line is computed excluding the HI–rich galaxies with $def_{\rm HI} < -0.3$ and the open squares indicates the profile for HI–deficient galaxies without NGC 4321. HI–deficient galaxies have in the inner part of the disk an HI column density large enough to form H_2.

At this point we have offered an interpretation for why the HI–deficiency is only a necessary condition for the in H_2–deficiency in HI poor galaxies, but a new question arises: if unperturbed and HI–deficient galaxies have a comparable gas column density, why do they have different values of $def_{\rm HI}$? The analysis presented in Figure 7.6 provides the answer. In this plot we show the average HI surface density profiles in bins of the normalized galactocentric radius up to twice R_{25}. As for Fig. 7.5, inspecting the individual HI profiles, we noticed that NGC 4321 exhibits an unusual high column density. To test if this is relevant for our analysis, we over-plot with open squares a new HI profile excluding this galaxy. This profile is no longer fully compatible with the one observed for non–deficient galaxies. It is also evident that the surface density profiles for HI–deficient and gas–deficient galaxies exhibit a different behaviour. The former rises towards the center, approaching the gas column density observed in unperturbed galaxies, while the latter remains at the lowest observed column density at all the radii. Assuming the profile for unperturbed galaxies serves as a template for how the atomic gas is distributed in spiral galaxies, HI–deficient galaxies have a significant residual quantity of gas towards the center that allows the H_2 formation, but have much less HI at larger radii. Therefore, the main contribution to $def_{\rm HI}$ comes from the outer part of the disk where the HI profile rapidly declines. Gas–deficient galaxies have an HI column density that is reduced relative to that of unperturbed galaxies in both their inner and outer disks and it is only this reduction in the HI content from the inner disk that gives rise to H_2 deficiency. The dashed line in Fig. 7.6 shows that the observed behaviour is not driven by HI rich galaxies.

In Figure 7.7 we show the ratio of the molecular to atomic hydrogen as a function of the total gas column density. In this plot, all systematic differences between the classes of galaxy disappear. Since we are exploring the high density regions, the fact that even gas-deficient galaxies show the same behaviour observed in non–perturbed and HI–deficient galaxies suggests that molecule formation is a local process within a galaxy and that the relevant local variable is

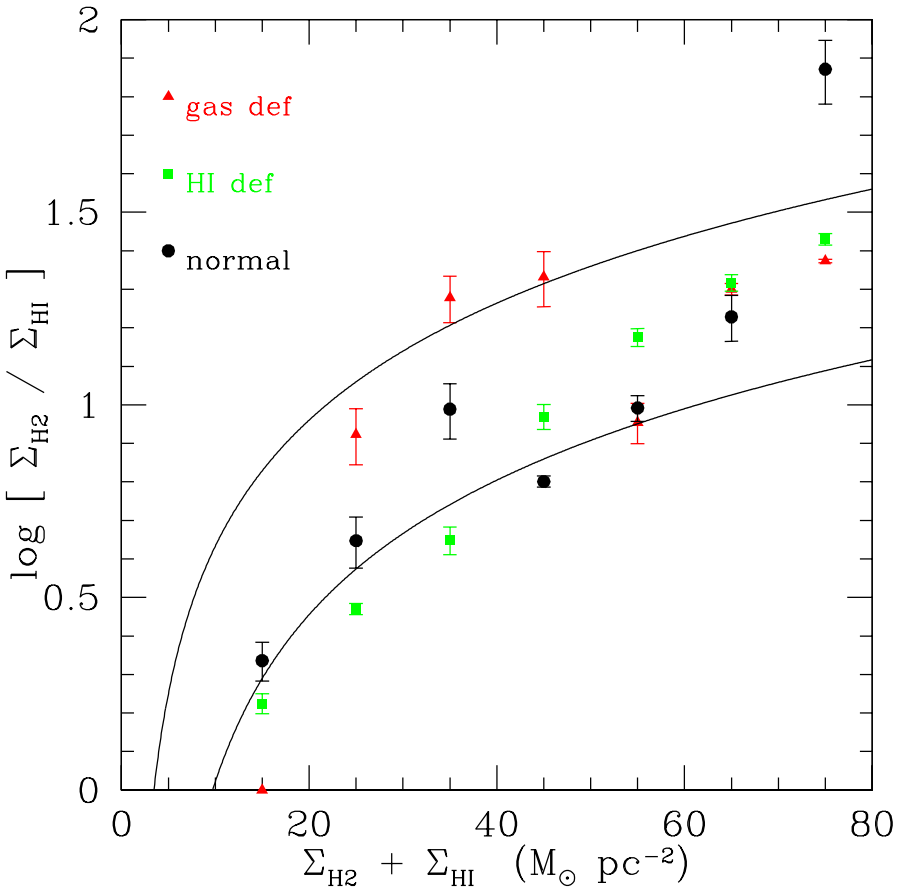

Figure 7.7: Ratio of the molecular to atomic hydrogen as a function of the total gas column density. The solid lines show the expected ratio according to the model by Krumholz et al. (2009a) for lower ($[O/H] = -0.4$) and upper ($[O/H] = 0$) value of metallicity in our sample. At high gas surface densities gas rich and gas poor galaxies exhibit the same behaviour because locally the physics of the H_2 formation remains unchanged.

column density. The solid lines in Fig. 7.7 are the expected values of molecular fraction from the model by Krumholz et al. (2009a). The model is dependent on the metallicity, so we consider a lower ($[O/H] = -0.4$) and an upper ($[O/H] = 0$) value that reasonably correspond to the range of metallicity observed in our sample. There are some approximations that limit the robustness of the comparison between the data and the model. First, the density considered in the model is computed assuming pressure balance, which may not hold in perturbed galaxies; second, we average the gas column density on physical scales that contain several atomic–molecular complexes, while the model is formulated for a single cloud. Despite all these limitations, there is a satisfactory agreement between data and the model. Finally, Figure 7.7 is useful to stress once again the main result of our analysis: galaxies which suffer from HI depletion inside the optical disk are on average shifted at lower column density where the production of H_2 becomes inefficient.

7.5.3 H_2–deficiency and star formation

Our analysis offers a possible solution to the current inconsistency in studies of the star formation and galaxy environment. In fact, the observed molecular deficiency is the missing link between the quenching of the star formation rate (SFR) in HI–deficient galaxies (e.g. Gavazzi et al. 2002) and the fact that the bulk of the stars form in molecular regions (Wong & Blitz 2002, Kennicutt et al. 2007, Fumagalli & Gavazzi 2008, Bigiel et al. 2008). Our simple picture is that, when perturbations reduce the atomic hydrogen column density below the threshold required to produce the molecular hydrogen, star formation will be suppressed.

We can test this hypothesis by comparing in Figure 7.8 the integrated atomic and molecular hydrogen content with our indicator for the star formation activity. The top–left panel reproduces the well–known trend (e.g. Gavazzi et al. 2002) between HI–deficiency and the quenching of star formation activity, measured through the $H\alpha + [NII]$ E.W.. Gas–deficient

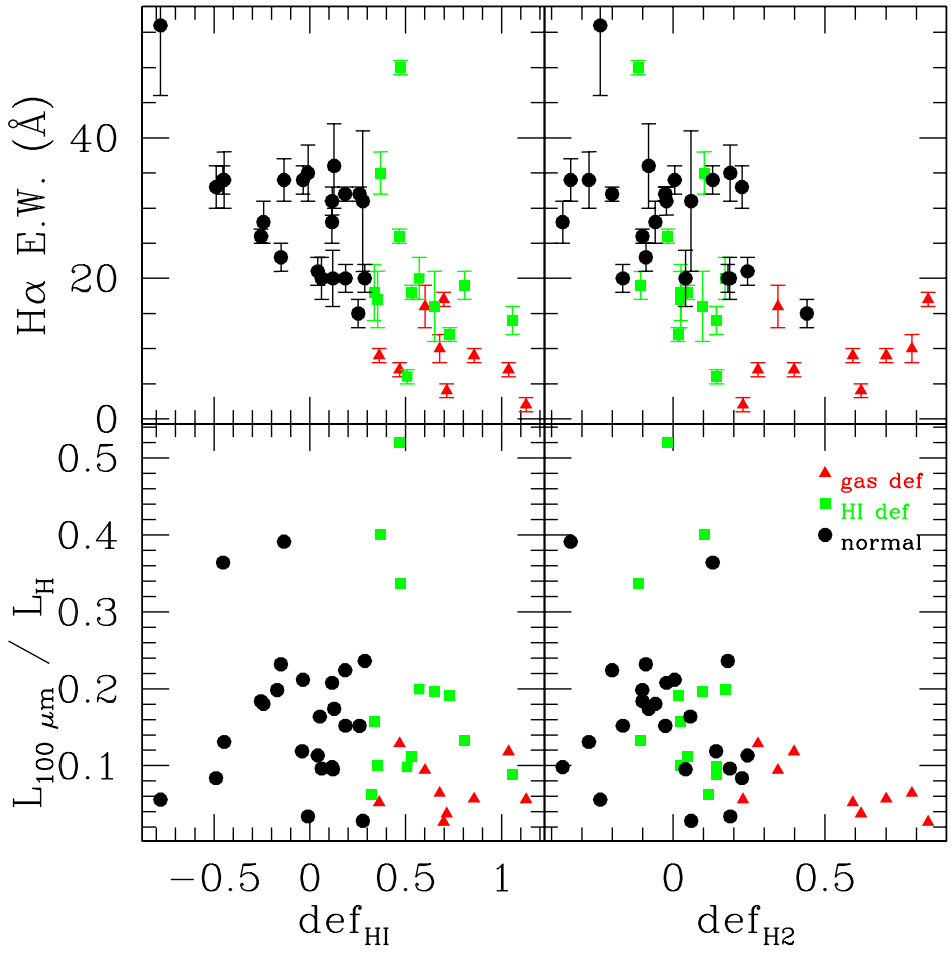

Figure 7.8: Comparison between the atomic and molecular gas deficiency with SFR indicators. The top panels show the $H\alpha+[NII]$ E.W., while the bottom panels show the 100μm luminosity normalised to the H–band luminosity. Gas–deficient galaxies have the lowest absolute specific star formation rate as expected if stars form in molecular rich regions.

galaxies (triangles) clearly exhibit a lower specific star formation rate than unperturbed (circles) and HI–deficient galaxies (squares). In the top–right panel, we plot the $H\alpha+[NII]$ E.W. versus $def_{\rm H_2}$. Two distinct behaviours can now be identified: molecular hydrogen deficient galaxies (triangles) have their star formation rate strongly suppressed with respect to molecular rich galaxies (squares and circles). Comparing both the top panels, we conclude that the trend between $def_{\rm HI}$ and the star formation rate is in reality a reflection of the link between $def_{\rm H_2}$ and $def_{\rm HI}$ with $def_{\rm H_2}$ and the star formation rate. In more detail, looking at the galaxies with $def_{\rm HI} \geq 0.3$ it appears that the HI–deficient galaxies can have star formation rates equal to the values observed in non–deficient galaxies, while gas–deficient galaxies have the lowest $H\alpha + [NII]$ E.W. The dispersion observed in the top–left panel can be interpreted as being primarily a product of the observed range in the molecular gas deficiency: not all HI–deficient galaxies are molecular deficient and thus not all them have suppressed star formation. This degeneracy is removed by directly comparing the $H\alpha+[NII]$ E.W. with the $def_{\rm H_2}$. Because we find the same result by comparing the 100μm luminosity normalized to the H–band luminosity with both $def_{\rm HI}$ and $def_{\rm H_2}$ (bottom panels), we are confident that dust extinction or $[NII]$ contamination do not significantly affect our conclusions.

7.6 Summary and conclusion

Using high sensitivity and high resolution CO images from the Nobeyama and the BIMA surveys we have studied molecular hydrogen deficiency in spiral galaxies. Our results and conclusions are summarised as follows.

– Comparing the total masses of molecular and atomic hydrogen, we detect H_2–deficiency in a subset of galaxies which are HI poor. The same result is found by studying directly the H_2 surface density profiles. We observe that molecular gas deficiency is associated

with the HI reduction inside the optical disk. The deficiency in the atomic hydrogen is a necessary condition for the molecular gas deficiency, but the $def_{\rm H_2}$ parameter spans a wide range of values even in galaxies which are HI–deficient. In fact, there are HI poor galaxies which have a normal H_2 content.

– When we examine the molecular fraction at a given point in a galaxy as a function of the total gas column density, we no longer detect any systematic difference between unperturbed, HI–deficient and gas–deficient galaxies. This suggests that the total gas column density, or something strongly correlated with it, is the primary factor in determining the molecular content of a galaxy. In this picture, a galaxy moves from HI–deficient to gas–deficient when it loses enough gas inside its optical disk to bring the gas column density below the critical value of ~ 10 $M_\odot$ pc^{-2} required for H_2 to be present. This behaviour is quantitatively consistent with the models of Krumholz et al. (2008; 2009a), albeit with significant uncertainties.

– Since the bulk of the star formation seems connected with the molecular gas, when perturbations reduce the atomic hydrogen column density below the threshold required to produce molecular hydrogen, star formation is suppressed. Studying the $H\alpha + [NII]$ E.W. or the 100μm luminosity, we show that H_2–deficient galaxies have less star formation than what is observed in H_2–normal galaxies and that the well–known trend between the HI–deficiency and the star formation rate most likely reflects a physical connection between molecular hydrogen depletion and star formation activity.

All together, these results corroborate a picture in which the gas phase as a whole feeds the star formation activity. The atomic hydrogen is the essential and primary fuel for the formation of molecular hydrogen that sustains the bulk of star formation activity. When environmental processes or feedback are able to reduce the atomic gas inside the optical disk,

the HI column density drops below the critical column density required for molecular hydrogen formation and star formation is quenched. This scenario is supported on cosmological scales, where the HI content in galaxies appears to be unchanged for the past $\sim$ 10 Gyr (Prochaska & Wolfe 2009). This suggests that the global star formation rate is driven by the accretion of fresh gas from the IGM. In addition, our analysis explains how environmental perturbations or galaxy feedback, which remove the atomic gas, can suppress the formation of the H_2 and stars. The behaviour seen in the innermost bin ($0 \leq R/R_{25} < 0.2$) of Figures 7.3, 7.5 and 7.6 allows to rule out the hypothesis that AGN feedback is the dominant process responsible for the gas depletion. In fact, gas is almost unperturbed in the galaxy center. This is expected in our sample where 7/9 of gas–deficient galaxies belong to the Virgo cluster or the Coma 1 cloud in which gas stripping is the most efficient process (Boselli & Gavazzi 2006, Boselli et al. 2006).

Ongoing CO surveys as HERACLES by Leroy et al. (in prep) or JCMT, the Nearby Galaxies Legacy Survey by Wilson et al. (2008), will soon provide larger samples to test our results. A preliminary comparison with the fluxes published by Wilson et al. (2008) appear to support our work. Looking at their Table 4, the four listed galaxies show from the top to the bottom decreasing values of SFR and CO $J: 3-2$ luminosity; consistently, we find for them an increasing value of $def_{\mathrm{H_2}}$. If confirmed, our results support the argument that the bulk of star formation is connected with the molecular gas phase. Moreover, the ability to derive the molecular gas content starting from an observed HI distribution will provide a useful way to correct the star formation rate expected from local star formation laws (Kennicutt 1998, Bigiel et al. 2008), including the effect of the molecular gas deficiency which suppresses star formation in rich clusters or in interacting galaxies.

Our analysis leaves open a fundamental question on the implications that H_2–deficiency might have for galaxy evolution. In the standard scenario where star formation is supported by cold gas infall towards the inner part of the galaxy halo (Boissier & Prantzos 1999, Dekel &

Birnboim 2006), it is fascinating to think that, when the gas flow is halted, the gas is devoured by ongoing star formation. The galaxy depletes its gaseous reservoirs and turns into a passive and quiescent system (van den Bergh 1976). This process might explain a first step for the transition between the blue and red sequence in the color–magnitude diagram. Interactions with the intergalactic medium might thus be an alternative process to AGN feedback or galaxy mergers to explain the departure of star forming systems from the blue sequence. In our scenario, two time–scales govern this migration, one set by the atomic gas depletion and the other related to the response of the star formation activity to the gas removal. The first varies according to the physical process responsible for the HI gas removal, but in this phase the galaxy is still forming stars (e.g. Boselli et al 2006). The second is a relatively short time–scale, at most set by the gas consumption through star formation ($\sim$ 2 Gyr, Bigiel et al. 2008)[2], but more likely connected with the much shorter life time of the GMCs. This second phase is consistent with a rapid evolutionary path in the color–magnitude diagram. However, although it has been proved that dwarf galaxies can turn into dwarf ellipticals within $\sim$ 2 Gyr in the local universe (Boselli et al. 2008), it is challenging to prove that massive spirals can turn into giant ellipticals just as a consequence of gas removal through interactions with the cluster IGM. In fact, their angular momentum would be fairly conserved in this kind of interaction. Major mergers as those dominating at higher redshifts must be invoked (Faber et al. 2007).

[2]This value comes from the ratio of gas surface density and star formation rate surface density which is assumed constant. However, gas recycling may increase Σ_{gas} up to 40% and the actual time–scale may be longer.

7.7 Appendix

7.7.1 HI fluxes

As discussed in Sec. 7.3.2.2, the HI fluxes quoted in the literature for individual galaxies are subject to discrepancies which in some case are not negligible. To show that our result is independent of a particular choice for the HI fluxes, we compute different sets of $def_{\rm HI}$ parameters using fluxes from ALFALFA, THINGS, VIVA and Springob et al. (2005). Figure 7.9 shows a comparison of these values with the HI–deficiency computed from HyperLeda fluxes adopted in our analysis.

Inspecting the left–panels, we notice that $def_{\rm HI}$ computed from ALFALFA and VIVA is in agreement with our values. Because these two surveys provide HI fluxes in the Virgo cluster where the most HI–deficient objects reside, we conclude that $def_{\rm HI}$ for these galaxies is not affected by important fluctuation. On the contrary, looking at the right–panels, we notice that at low HI–deficiencies a systematic offset appears. Since only few galaxies at low deficiency are available in ALFALFA or VIVA, an additional comparison is almost impossible and it is difficult to further discuss this difference. However, assuming that HyperLeda fluxes suffer from a systematic offset, we should correct our $def_{\rm HI}$ towards even lower $def_{\rm HI}$ values and galaxies we classify as non–deficient remain in this class. This is not the case for NGC 5236 and NGC 6946, the two outliers visible in the bottom–right panel. However, these galaxies are more extended than the primary beam mapped by THINGS and we are confident that our value for $def_{\rm HI}$ is more accurate.

This comparison confirms that, despite the fluctuation in the $def_{\rm HI}$ parameter, the final distribution for the HI–deficiency and thus our analysis are not significantly affected by a particular choice of the HI flux.

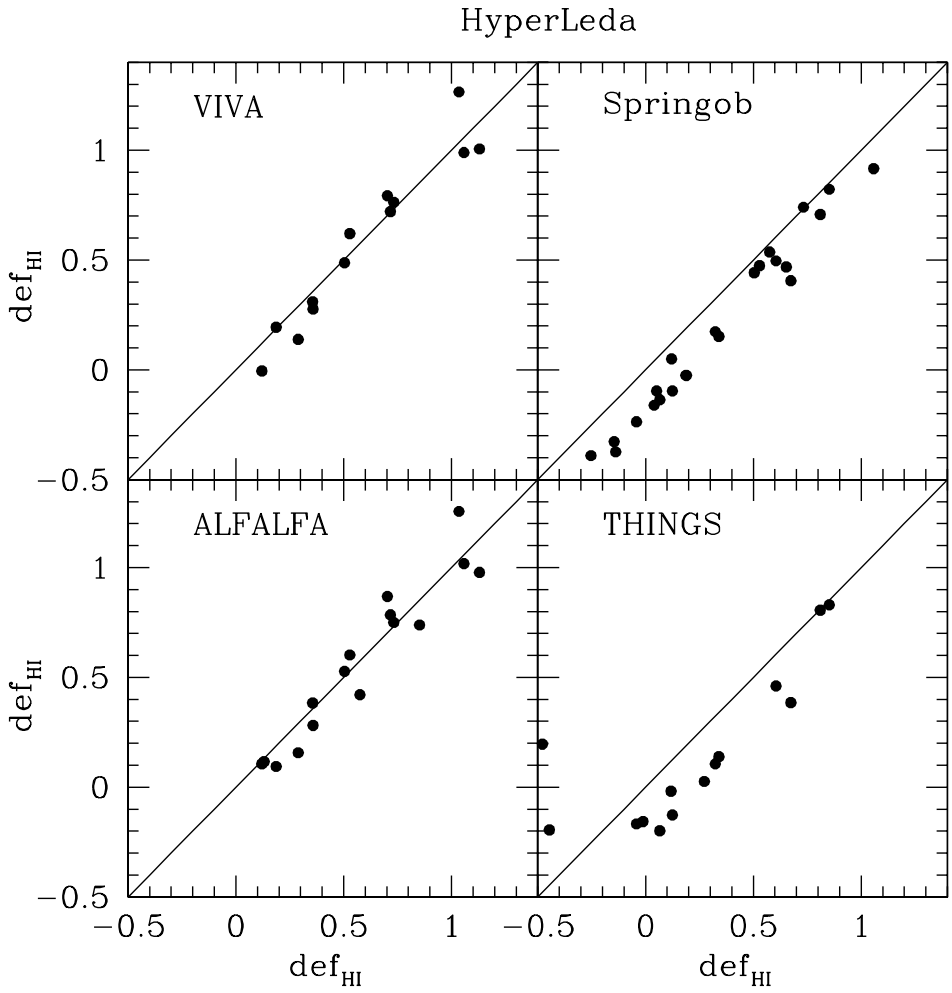

Figure 7.9: Comparison between the HI–deficiency computed using fluxes from HyperLeda (x–axis) and from THINGS, VIVA, ALFALFA and Springob et al. (2005) (y–axis). The solid line is the 1:1 correlation.

Chapter 8

Summary and future directions

In this thesis we have examined two of the key processes in the assembly and evolution of high-redshift galaxies: gas accretion from the intergalactic medium and the conversion of gas into stars.

In Chapter 2, from Fumagalli et al. (2011a), we used hydrodynamic simulations of galaxy formation to investigate what observational signatures can be used to probe gas accretion onto galaxies in the form of filamentary streams. Our principal finding is that absorption lines, and in particular the study of (metal poor) Lyman limit systems, are the most suitable probes to uncover with observations whether the predicted mode of accretion operates in real galaxies.

These theoretical predictions are complemented by new observations presented in Chapter 3 where we analyzed the chemical and physical properties of two Lyman limit systems at $z \sim 3$ and $z \sim 3.5$ that are composed by cold gas with no discernible metal lines (Fumagalli et al. 2011b). These systems not only provide a robust empirical verification of Big Bang nucleosynthesis, but also are compelling candidates of cold streams, the mode of gas accretion currently predicted by numerical simulations. This study offers a proof-of-concept that the analysis of a large sample of Lyman limit systems will play a key role in constraining models

of galaxy formation in the near future.

Starting from Chapter 4, we turned our attention to the second key ingredient for the assembly and evolution of galaxies: the conversion of gas into stars. In Chapter 4 and 5 we presented results from a new imaging survey of 32 quasar fields that host damped Lyman$-\alpha$ systems (Fumagalli et al. 2010b; 2012). Since the necessary condition for the detection of damped Lyman$-\alpha$ systems is the presence of a bright background quasar, previous studies that attempted to connect directly the gas absorption to the galaxy emission were often plagued by the glare of the background sources. With a new technique that overcomes this limitation (O'Meara et al. 2006), we were able to probe for the first time the relationship between the hydrogen column density and the *in situ* star formation in damped Lyman$-\alpha$ systems. Our principal finding is that damped Lyman$-\alpha$ galaxies do not form stars at appreciable rates ($\gtrsim$ $1 M_\odot$ yr^{-1}) either in compact regions (3 kpc) or in more extended disks (10 kpc), corresponding to local star formation surface densities of $< 0.01-0.03 M_\odot$ yr^{-1} kpc^{-2}. Further, the distribution of galaxies in the surroundings of these quasar sightlines reveals that typical damped Lyman$-\alpha$ systems do not arise from extended disks ($0 - 10$ kpc) around $\gtrsim 0.1 L_*$ galaxies.

Finally, Chapter 6 and Chapter 7 (Fumagalli et al. 2009; 2010a) offered the theoretical framework to interpret the lack of *in situ* star formation in damped Lyman$-\alpha$ galaxies. Using multiwavelength imaging of nearby spiral and irregular galaxies, we tested models for the conversion of atomic to molecular hydrogen, a process that closely traces the thermodynamic conditions that are the most favorable for star formation. Through this analysis, we verified that the formation of molecular hydrogen locally depends on the hydrogen column density and metallicity of the gas. At low column density and low metallicity, gas cannot effectively shield from the ambient radiation field and the ability to cool and form stars is suppressed. We attribute the dearth of star formation in damped Lyman$-\alpha$ systems to these processes.

While this thesis provides additional clues to solve the riddles associated to the ac-

cretion of gas onto distant galaxies and the conversion of this newly acquired fuel into stars, this work clearly does not resolve the problem of developing a coherent theory for the baryon processes inside forming galaxies. Progress can now be made in three different directions.

First, the theoretical predictions presented in this work provide only general guidelines for the observables that can be used to probe accretion in distant galaxies. A more quantitative description of the different statistics such as the covering factor of optically thick gas or the distribution of metals in the circumgalactic medium of the simulated galaxies is now desired. In particular, it should be extensively investigated to what degree these predictions are robust with respect to different numerical codes, different radiative transfer techniques, and the different sub-grid prescriptions for star formation and feedback. At the same time, high-quality spectroscopy for a large sample of Lyman limit systems is now desired to map with higher fidelity the shape and redshift evolution of the hydrogen and metallicity distribution function in this class of absorbers. The overarching goal is to produce trustworthy statistics from simulations and observations that can be used to finally learn about the mechanisms by which galaxies acquire their gas. The first steps in this direction are being made.

Second, there should be a continued effort to connect individual absorbers to their galaxy counterparts. The statistical analysis of absorption line systems already provides a wealth of information, but the possibility to build larger samples of galaxy/absorber pairs at high redshifts is an invaluable opportunity to learn about the evolution of galaxies in relation to the amount of gas available (e.g. Tumlinson et al. 2011). Large samples of damped Lyman-α galaxies for which the emitting counterparts are known will, for example, allow one to resolve the ambiguity found in this study: do the majority of damped Lyman-α systems arise from faint galaxies that are centrals to their halos or do satellites in the surroundings of more massive halos play a significant role? The answer to this question has profound implication to understand the cosmic evolution of neutral hydrogen (Prochaska & Wolfe 2009). Further, the ability to con-

struct samples of Lyman limit systems that are associated with known galaxies will be a crucial observational test for whether most of the Lyman limit systems arise from the circumgalactic medium of actively star forming galaxies. In the next few years, large field-of-view integral field unit spectrographs or infrared multiobject spectrographs will be key instruments for this type of study.

Third, in this thesis there has been only a limited discussion about feedback, the third key ingredient for a complete theory of galaxy formation. However, a more detailed understanding of the complex interplay between gas accretion and galactic winds in galaxies becomes necessary at this point. This is true from a theoretical point of view because results of numerical simulations are highly sensitive to the assumed feedback schemes (Scannapieco et al. 2012). At the same time, without any doubt, galactic winds play also a role in shaping the observed properties of absorption line systems. To quantify these effects is a challenging task that has to be undertaken.

In light of the discussion presented in this thesis, we conclude with a positive outlook over the future of galaxy formation studies. We believe in fact that in the coming years there are exciting prospects to finally formulate a coherent theory for the baryon processes in distant galaxies.

Bibliography

Abazajian, K. N., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2009, ApJS, 182, 543

Abdo, A. A., Ackermann, M., Ajello, M., et al. 2010, A&A, 523, L2

Abel, T., & Wandelt, B. D. 2002, MNRAS, 330, L53

Adelberger, K. L., Steidel, C. C., Shapley, A. E., & Pettini, M. 2003, ApJ, 584, 45

Adelman-McCarthy, J. K., et al. 2008, ApJS, 175, 297

Agertz, O., Teyssier, R., & Moore, B. 2009, MNRAS, 397, L64

Aguirre, A., Dow-Hygelund, C., Schaye, J., & Theuns, T. 2008, ApJ, 689, 851

Albrecht, M., Chini, R., Krügel, E., Müller, S. A. H., & Lemke, R. 2004, A&A, 414, 141

Altay, G., Theuns, T., Schaye, J., Crighton, N. H. M., & Dalla Vecchia, C. 2010, arXiv:1012.4014

Aragon-Salamanca A., Ellis R. S., O'Brien K. S., 1996, MNRAS, 281, 945

Asplund, M., Grevesse, N., Sauval, A. J., & Scott, P. 2009, ARA&A, 47, 481

Bahcall, J. N., & Peebles, P. J. E. 1969, ApJ, 156, L7

Barnes L. A., Haehnelt M. G., 2009, MNRAS, 397, 511

Barone, L. T., Heithausen, A., Hüttemeister, S., Fritz, T., & Klein, U. 2000, MNRAS, 317, 649

Beck, S. C., Turner, J. L., Langland-Shula, L. E., Meier, D. S., Crosthwaite, L. P., & Gorjian, V. 2002, AJ, 124, 2516

Bell, E. F., & de Jong, R. S. 2001, ApJ, 550, 212

Bell, E. F., McIntosh, D. H., Katz, N., & Weinberg, M. D. 2003, ApJS, 149, 289

Bertin E., 2009, Memorie della Societa Astronomica Italiana, 80, 422

Bertin, E. 2006, Astronomical Data Analysis Software and Systems XV, 351, 112

Bertin, E., Mellier, Y., Radovich, M., et al. 2002, Astronomical Data Analysis Software and Systems XI, 281, 228

Bertin E., Arnouts S., 1996, A&AS, 117, 393

Bigiel, F., Leroy, A., Walter, F., Brinks, E., de Blok, W. J. G., Madore, B., & Thornley, M. D. 2008, AJ, 136, 2846

Bigiel, F., Leroy, A., Walter, F., et al. 2010, AJ, 140, 1194

Birnboim, Y., & Dekel, A. 2003, MNRAS, 345, 349

Birnboim, Y., & Dekel, A. 2010, arXiv:1008.1060

Blitz, L. & Rosolowsky, E. 2004, ApJ, 612, L29

Blitz, L., & Rosolowsky, E. 2006, ApJ, 650, 933

Bloemen, J. B. G. M., Strong, A. W., Mayer-Hasselwander, H. A., et al. 1986, A&A, 154, 25

Boissier, S., & Prantzos, N. 1999, MNRAS, 307, 857

Boissier, S., Prantzos, N., Boselli, A., & Gavazzi, G. 2003, MNRAS, 346, 1215

Bolatto, A. D., Leroy, A. K., Rosolowsky, E., Walter, F., & Blitz, L. 2008, ApJ, 686, 948

Bolatto, A. D., Leroy, A. K., Jameson, K., et al. 2011, ApJ, 741, 12

Bolton A. S., Burles S., Koopmans L. V. E. et al., 2006, ApJ, 638, 703

Boselli, A., & Gavazzi, G. 2006, PASP, 118, 517

Boselli, A., Gavazzi, G., Lequeux, J., Buat, V., Casoli, F., Dickey, J., & Donas, J. 1997, A&A, 327, 522

Boselli, A., Gavazzi, G., Donas, J., & Scodeggio, M. 2001, AJ, 121, 753

Boselli, A., Lequeux, J., & Gavazzi, G. 2002, A&A, 384, 33

Boselli, A., Boissier, S., Cortese, L., Gil de Paz, A., Seibert, M., Madore, B. F., Buat, V., & Martin, D. C. 2006, ApJ, 651, 811

Boselli, A., Boissier, S., Cortese, L., & Gavazzi, G. 2008, ApJ, 674, 742

Bothwell, M. S., Kennicutt, R. C., & Lee, J. C. 2009, MNRAS, 400, 154

Bouché, N., Dekel, A., Genzel, R., et al. 2010, ApJ, 718, 1001

Bouwens R. J., Illingworth G. D., Blakeslee J. P. et al., 2004, ApJ, 611, L1

Bowen, D. V., Jenkins, E. B., Pettini, M., & Tripp, T. M. 2005, ApJ, 635, 880

Brammer G. B., van Dokkum P. G., Coppi P., 2008, ApJ, 686, 1503

Bromm, V., Ferrara, A., Coppi, P. S., & Larson, R. B. 2001, MNRAS, 328, 969

Brooks, A. M., Governato, F., Quinn, T., Brook, C. B., & Wadsley, J. 2009, ApJ, 694, 396

Bunker A. J., Warren S. J., Clements D. L. et al., 1999, MNRAS, 309, 875

Burbidge, E. M., Burbidge, G. R., Fowler, W. A., & Hoyle, F. 1957, Reviews of Modern Physics, 29, 547

Burles, S., & Tytler, D. 1998, ApJ, 507, 732

Burles, S., & Tytler, D. 1998, ApJ, 499, 699

Caffau, E., Bonifacio, P., François, P., et al. 2011, Nature, 477, 67

Caldwell, N., & Phillips, M. M. 1989, ApJ, 338, 789

Calzetti, D., Meurer, G. R., Bohlin, R. C., Garnett, D. R., Kinney, A. L., Leitherer, C., & Storchi-Bergmann, T. 1997, AJ, 114, 1834

Campos-Aguilar, A., Moles, M., & Masegosa, J. 1993, AJ, 106, 1784

Cantalupo, S. 2010, MNRAS, 403, L16

Cantalupo, S., & Porciani, C. 2011, MNRAS, 411, 1678

Cantalupo, S., Porciani, C., Lilly, S. J., & Miniati, F. 2005, ApJ, 628, 61

Cardelli J. A., Clayton G. C., Mathis J. S., 1989, ApJ, 345, 245

Casoli, F., Boisse, P., Combes, F., & Dupraz, C. 1991, A&A, 249, 359

Cen, R. 2012, ApJ, 748, 121

Ceverino, D., & Klypin, A. 2009, ApJ, 695, 292

Ceverino, D., Dekel, A., & Bournaud, F. 2010, MNRAS, 404, 2151

Chen H., Helsby J. E., Gauthier J. et al., 2010, ArXiv e-prints

Christensen, L., Wisotzki, L., Roth, M. M., Sánchez, S. F., Kelz, A., & Jahnke, K. 2007, A&A, 468, 587

Christensen L., Noterdaeme P., Petitjean P. , 2009, A&A, 505, 1007

Chung, A., van Gorkom, J. H., Kenney, J. D. P., Crowl, H., & Vollmer, B. 2009, AJ, 138, 1741

Ciardullo, R., Gronwall, C., Wolf, C., et al. 2012, ApJ, 744, 110

Cooke J., Wolfe A. M., Prochaska J. X. et al., 2005, ApJ, 621, 596

Cooke J., Wolfe A. M., Gawiser E. et al., 2006, ApJ, 636, L9

Crighton, N. H. M., Webb, J. K., Ortiz-Gil, A., & Fernández-Soto, A. 2004, MNRAS, 355, 1042

Croton, D. J., Springel, V., White, S. D. M., et al. 2006, MNRAS, 365, 11

Daddi, E., et al. 2010, ApJ, 713, 686

Dale, D. A., Helou, G., Neugebauer, G., Soifer, B. T., Frayer, D. T., & Condon, J. J. 2001, AJ, 122, 1736

Dall'Aglio, A., Wisotzki, L., & Worseck, G. 2009, arXiv:0906.1484

Davé, R., Finlator, K., & Oppenheimer, B. D. 2012, MNRAS, 421, 98

Davé, R., Oppenheimer, B. D., & Finlator, K. 2011, MNRAS, 415, 11

Davidge, T. J. 1989, PASP, 101, 494

Decarli, R., Gavazzi, G., Arosio, I., Cortese, L., Boselli, A., Bonfanti, C., & Colpi, M. 2007, MNRAS, 381, 136

de Grijs, R., et al. 2004, MNRAS, 352, 263

De Lucia, G., & Blaizot, J. 2007, MNRAS, 375, 2

Dekel, A., & Birnboim, Y. 2006, MNRAS, 368, 2

Dekel, A., & Birnboim, Y. 2008, MNRAS, 383, 119

Dekel, A., Sari, R., & Ceverino, D. 2009, ApJ, 703, 785

Dekel, A., et al. 2009, Nature, 457, 451

Dessauges-Zavadsky M., Calura F., Prochaska J. X. et al., 2007, A&A, 470, 431

de Vaucouleurs, G., de Vaucouleurs, A., Corwin, H. G., Jr., Buta, R. J., Paturel, G., & Fouque, P. 1991, Volume 1-3, XII, 2069 pp. 7 figs.. Springer-Verlag Berlin Heidelberg New York,

Dickman, R. L., Snell, R. L., & Schloerb, F. P. 1986, ApJ, 309, 326

Dijkstra, M., & Loeb, A. 2009, MNRAS, 400, 1109

Dijkstra, M., & Hultman Kramer, R. 2012, arXiv:1203.3803

di Serego Alighieri, S., et al. 2007, A&A, 474, 851

Djorgovski S. G., Pahre M. A., Bechtold J. et al., 1996, Nature, 382, 234

D'Odorico, V., & Petitjean, P. 2001, A&A, 370, 729

Downes A. J. B., Peacock J. A., Savage A. et al., 1986, MNRAS, 218, 31

Draine, B. T. 2003, ARA&A, 41, 241

Drissen, L., Roy, J.-R., Robert, C., Devost, D., & Doyon, R. 2000, AJ, 119, 688

Dwek, E. 1998, ApJ, 501, 643

Ekta, B., Pustilnik, S. A., & Chengalur, J. N. 2009, MNRAS, 397, 963

Ellison S. L., Pettini M., Steidel C. C. et al., 2001, ApJ, 549, 770

Ellison, S. L., Kanekar, N., Prochaska, J. X., Momjian, E., & Worseck, G. 2012, arXiv:1205.1200

Elmegreen, B. G. 1989, ApJ, 338, 178

Elmegreen, B. G. 1993, ApJ, 411, 170

Erb, D. K., Shapley, A. E., Pettini, M., et al. 2006, ApJ, 644, 813

Fabbian, D., Nissen, P. E., Asplund, M., Pettini, M., & Akerman, C. 2009, A&A, 500, 1143

Faber, S. M., et al. 2007, ApJ, 665, 265

Fardal, M. A., Katz, N., Gardner, J. P., Hernquist, L., Weinberg, D. H., & Davé, R. 2001, ApJ, 562, 605

Faucher-Giguère, C.-A., & Kereš, D. 2011, MNRAS, 412, L118

Faucher-Giguère, C.-A., Lidz, A., Hernquist, L., & Zaldarriaga, M. 2008, ApJ, 682, L9

Faucher-Giguère, C.-A., Kereš, D., Dijkstra, M., Hernquist, L., & Zaldarriaga, M. 2010, ApJ, 725, 633

Faucher-Giguère, C. -., Keres, D., & Ma, C. -. 2011, arXiv:1103.0001

Feldmann, R., Gnedin, N. Y., & Kravtsov, A. V. 2012, ApJ, 747, 124

Ferguson H. C., Dickinson M., Giavalisco M. et al., 2004, ApJ, 600, L107

Ferland, G. J., Korista, K. T., Verner, D. A., Ferguson, J. W., Kingdon, J. B., & Verner, E. M. 1998, PASP, 110, 761

Ferrara, A., Pettini, M., & Shchekinov, Y. 2000, MNRAS, 319, 539

Förster Schreiber N. M., Genzel R., Bouché N. et al., 2009, ApJ, 706, 1364

Fox A. J., Prochaska J. X., Ledoux C. et al., 2009, A&A, 503, 731

Fumagalli, M., & Gavazzi, G. 2008, A&A, 490, 571

Fumagalli, M., Krumholz, M. R., Prochaska, J. X., Gavazzi, G., & Boselli, A. 2009, ApJ, 697, 1811

Fumagalli, M., Krumholz, M. R., & Hunt, L. K. 2010, ApJ, 722, 919

Fumagalli, M.,O'Meara, J. M., Prochaska, J. X., & Kanekar, N. 2010, MNRAS, 408, 362

Fumagalli, M., Prochaska, J. X., Kasen, D., et al. 2011, MNRAS, 418, 1796

Fumagalli, M., O'Meara, J. M., & Prochaska, J. X. 2011, Science, 334, 1245

Fumagalli, M., et al. 2012, in preparation

Furlanetto, S. R., Schaye, J., Springel, V., & Hernquist, L. 2005, ApJ, 622, 7

Fynbo J. U., Møller P., Warren S. J., 1999, MNRAS, 305, 849

Fynbo, J. P. U., Ledoux, C., Noterdaeme, P., et al. 2011, MNRAS, 413, 2481

Fynbo, J. P. U., Laursen, P., Ledoux, C., et al. 2010, MNRAS, 408, 2128

Gallazzi, A., et al. 2009, ApJ, 690, 1883

Gao, Y., & Solomon, P. M. 2004, ApJ, 606, 271

Gavazzi, G., Boselli, A., Pedotti, P., Gallazzi, A., & Carrasco, L. 2002, A&A, 396, 449

Gavazzi, G., Boselli, A., Donati, A., Franzetti, P., & Scodeggio, M. 2003, A&A, 400, 451

Gavazzi, G., et al. 2008, A&A, 482, 43

Genzel R., Tacconi L. J., Eisenhauer, F. et al., 2006, Nature, 442, 786

Genzel, R., Tacconi, L. J., Gracia-Carpio, J., et al. 2010, MNRAS, 407, 2091

Giavalisco, M., et al. 2011, arXiv:1106.1205

Gil de Paz, A., Silich, S. A., Madore, B. F., Sánchez Contreras, C., Zamorano, J., & Gallego, J. 2002, ApJ, 573, L101

Gill, S. P. D., Knebe, A., & Gibson, B. K. 2004, MNRAS, 351, 399

Gilmour R., Gray M. E., Almaini O. et al., 2007, MNRAS, 380, 1467

Giovanelli, R., & Haynes, M. P. 1985, ApJ, 292, 404

Giovanelli, R., et al. 2005, AJ, 130, 2598

Glover, S. C. O., & Clark, P. C. 2012, MNRAS, 421, 9

Gnat, O., & Sternberg, A. 2007, ApJS, 168, 213

Gnedin, N. Y. 2010, ApJ, 721, L79

Gnedin, N. Y., & Abel, T. 2001, New Astronomy, 6, 437

Gnedin, N. Y., & Kravtsov, A. V. 2010, ApJ, 714, 287

Gnedin, N. Y., Tassis, K., & Kravtsov, A. V. 2009, ApJ, 697, 55

Goerdt, T., Dekel, A., Sternberg, A., Ceverino, D., Teyssier, R., & Primack, J. R. 2010, MNRAS, 407, 613

Governato F., Willman B., Mayer L. et al., 2007, MNRAS, 374, 1479

Governato, F., Brook, C., Mayer, L., et al. 2010, Nature, 463, 203

Grazian A., Fontana A., de Santis C. et al., 2006, A&A, 449, 951

Grazian A., Menci N., Giallongo E. et al, 2009, A&A, 505, 1041

Guedes, J., Callegari, S., Madau, P., & Mayer, L. 2011, ApJ, 742, 76

Haardt, F., & Madau, P. 1996, ApJ, 461, 20

Haardt, F., & Madau, P. 2011, arXiv:1103.5226

Haehnelt, M. G., Steinmetz, M., & Rauch, M. 1998, ApJ, 495, 647

Haiman, Z., Spaans, M., & Quataert, E. 2000, ApJ, 537, L5

Hayes, M., Scarlata, C., & Siana, B. 2011, Nature, 476, 304

Haynes, M. P., & Giovanelli, R. 1984, AJ, 89, 758

Helfer, T. T., Thornley, M. D., Regan, M. W., Wong, T., Sheth, K., Vogel, S. N., Blitz, L., & Bock, D. C.-J. 2003, ApJS, 145, 259

Hennawi, J. F., & Prochaska, J. X. 2007, ApJ, 655, 735

Hennawi, J. F., et al. 2006, ApJ, 651, 61

Hernquist, L., & Springel, V. 2003, MNRAS, 341, 1253

Hildebrandt H., Wolf C., Benítez N., 2008, A&A, 480, 703

Hirashita, H., & Hunt, L. K. 2004, A&A, 421, 555

Hong, S., Katz, N., Davé, R., Fardal, M., Kereš, D., & Oppenheimer, B. D. 2010, arXiv:1008.4242

Hopkins, A. M., Schulte-Ladbeck, R. E., & Drozdovsky, I. O. 2002, AJ, 124, 862

Hopkins, P. F., Murray, N., Quataert, E., & Thompson, T. A. 2010, MNRAS, 401, L19

Hunt, L. K., Vanzi, L., & Thuan, T. X. 2001, A&A, 377, 66

Hunt, L. K., Giovanardi, C., & Helou, G. 2002, A&A, 394, 873

Hunt, L. K., Thuan, T. X., & Izotov, Y. I. 2003, ApJ, 588, 281

Hunt, L. K., Dyer, K. K., Thuan, T. X., & Ulvestad, J. S. 2004, ApJ, 606, 853

Hunt, L., Bianchi, S., & Maiolino, R. 2005a, A&A, 434, 849

Hunt, L. K., Dyer, K. K., & Thuan, T. X. 2005b, A&A, 436, 837

Hunter, D. A., van Woerden, H., & Gallagher, J. S., III 1994a, ApJS, 91, 79

Hunter, D. A., O'Connell, R. W., & Gallagher, J. S., III 1994b, AJ, 108, 84

Imara, N., & Blitz, L. 2007, ApJ, 662, 969

Izotov, Y. I., & Thuan, T. X. 2004a, ApJ, 602, 200

Izotov, Y. I., & Thuan, T. X. 2004b, ApJ, 616, 768

Izotov, Y. I., Thuan, T. X., & Lipovetsky, V. A. 1997, ApJS, 108, 1

James, P. A., et al. 2004, VizieR Online Data Catalog, 341, 40023

Jedamzik K., Prochaska J. X., 1998, MNRAS, 296, 430

Jorgenson, R. A., Wolfe, A. M., & Prochaska, J. X. 2010, ApJ, 722, 460

Kacprzak G. G., Churchill C. W., Ceverino D. et al., 2009, ArXiv e-prints

Kacprzak, G. G., Churchill, C. W., Ceverino, D., Steidel, C. C., Klypin, A., & Murphy, M. T. 2010, ApJ, 711, 533

Kanekar N., Chengalur J. N., 2003, A&A, 399, 857

Kanekar N., Smette A., Briggs F. H. et al., 2009, ApJ, 705, L40

Kasen, D., et al. 2012, ApJ, in preparation

Katz, N., Hernquist, L., & Weinberg, D. H. 1992, ApJ, 399, L109

Kauffmann, G., White, S. D. M., & Guiderdoni, B. 1993, MNRAS, 264, 201

Kenney, J. D., & Young, J. S. 1986, ApJ, 301, L13

Kenney, J. D., & Young, J. S. 1988, ApJS, 66, 261

Kenney, J. D. P., & Young, J. S. 1989, ApJ, 344, 171

Kennicutt, R. C., Jr. 1983, AJ, 88, 483

Kennicutt, R. C., Jr. 1983, ApJ, 272, 54

Kennicutt, R. C., Jr. 1992, ApJ, 388, 310

Kennicutt, R. C., Jr. 1992, ApJ, 388, 310

Kennicutt, R. C., Jr. 1998, ApJ, 498, 541

Kennicutt, R. C., Jr., & Kent, S. M. 1983, AJ, 88, 1094

Kennicutt, R. C., Jr., et al. 2007, ApJ, 671, 333

Kennicutt, R. C., Jr., Lee, J. C., Funes, S. J., José G., Sakai, S., & Akiyama, S. 2008, ApJS, 178, 247

Kennicutt, R. C., Jr., & Evans, N. J., II 2012, arXiv:1204.3552

Kereš, D., & Hernquist, L. 2009, ApJ, 700, L1

Kereš, D., Katz, N., Weinberg, D. H., & Davé, R. 2005, MNRAS, 363, 2

Kereš, D., Katz, N., Fardal, M., Davé, R., & Weinberg, D. H. 2009, MNRAS, 395, 160

Kim T., Carswell R. F., Cristiani S. et al., 2002, MNRAS, 335, 555

Kimm, T., Slyz, A., Devriendt, J., & Pichon, C. 2011, MNRAS, 413, L51

Kirkman, D., Tytler, D., Burles, S., Lubin, D., & O'Meara, J. M. 2000, ApJ, 529, 655

Kirkman, D., Tytler, D., Suzuki, N., O'Meara, J. M., & Lubin, D. 2003, ApJS, 149, 1

Klypin, A. A., Trujillo-Gomez, S., & Primack, J. 2011, ApJ, 740, 102

Knollmann, S. R., & Knebe, A. 2009, ApJS, 182, 608

Kobulnicky, H. A., & Skillman, E. D. 2008, AJ, 135, 527

Kohler, K., & Gnedin, N. Y. 2007, ApJ, 655, 685

Komatsu, E., et al. 2009, ApJS, 180, 330

Komatsu, E., Smith, K. M., Dunkley, J., et al. 2011, ApJS, 192, 18

Koopmann, R. A., & Kenney, J. D. P. 2004, ApJ, 613, 866

Kravtsov, A. V. 2003, ApJ, 590, L1

Kravtsov, A. V., Klypin, A. A., & Khokhlov, A. M. 1997, ApJS, 111, 73

Krogager, J.-K., Fynbo, J. P. U., Møller, P., et al. 2012, arXiv:1204.2833

Kroupa, P. 2001, MNRAS, 322, 231

Krumholz, M. R., & McKee, C. F. 2005, ApJ, 630, 250

Krumholz, M. R., & Gnedin, N. Y. 2011, ApJ, 729, 36

Krumholz, M. R., McKee, C. F., & Tumlinson, J. 2008, ApJ, 689, 865

Krumholz, M. R., McKee, C. F., & Tumlinson, J. 2009, ApJ, 693, 216

Krumholz, M. R., McKee, C. F., & Tumlinson, J. 2009, ApJ, 699, 850

Krumholz, M. R., Ellison, S. L., Prochaska, J. X., & Tumlinson, J. 2009, ApJ, 701, L12

Krumholz, M. R., Leroy, A. K., & McKee, C. F. 2011, ApJ, 731, 25

Kulkarni V. P., Hill J. M., Schneider G. et al., 2000, ApJ, 536, 36

Kulkarni V. P., Hill J. M., Schneider G. et al., 2001, ApJ, 551, 37

Kulkarni V. P., Woodgate B. E., York D. G. et al., 2006, ApJ, 636, 30

Kuno, N., et al. 2007, PASJ, 59, 117

Kunth, D., Lequeux, J., Sargent, W. L. W., & Viallefond, F. 1994, A&A, 282, 709

Landolt A. U., 1992, AJ, 104, 340

Larson, R. B. 1981, MNRAS, 194, 809

Laursen, P., Sommer-Larsen, J., & Andersen, A. C. 2009, ApJ, 704, 1640

Lebouteiller, V., Kunth, D., Lequeux, J., Lecavelier des Etangs, A., Désert, J.-M., Hébrard, G., & Vidal-Madjar, A. 2004, A&A, 415, 55

Lebouteiller, V., Kunth, D., Thuan, T. X., & Désert, J. M. 2009, A&A, 494, 915

Lecavelier des Etangs, A., Désert, J.-M., Kunth, D., Vidal-Madjar, A., Callejo, G., Ferlet, R., Hébrard, G., & Lebouteiller, V. 2004, A&A, 413, 131

Lee, J. C., Salzer, J. J., Impey, C., Thuan, T. X., & Gronwall, C. 2002, AJ, 124, 3088

Lee, H., Skillman, E. D., Cannon, J. M., Jackson, D. C., Gehrz, R. D., Polomski, E. F., & Woodward, C. E. 2006, ApJ, 647, 970

Leibundgut B., Robertson J. G., 1999, MNRAS, 303, 711

Leitherer, C., et al. 1999, ApJS, 123, 3

Lejeune, T., & Schaerer, D. 2001, A&A, 366, 538

Leon, S., Combes, F., & Menon, T. K. 1998, A&A, 330, 37

Leroy, A., Bolatto, A. D., Simon, J. D., & Blitz, L. 2005, ApJ, 625, 763

Leroy, A., Cannon, J., Walter, F., Bolatto, A., & Weiss, A. 2007, ApJ, 663, 990

Leroy, A. K., Walter, F., Brinks, E., Bigiel, F., de Blok, W. J. G., Madore, B., & Thornley, M. D. 2008, AJ, 136, 2782

Leroy, A. K., et al. 2009, ApJ, 702, 352

Leroy, A. K., Walter, F., Bigiel, F., et al. 2009, AJ, 137, 4670

Leroy, A. K., Bolatto, A., Gordon, K., et al. 2011, ApJ, 737, 12

Levshakov, S. A., Agafonova, I. I., Centurión, M., & Molaro, P. 2003, A&A, 397, 851

Levshakov, S. A., Agafonova, I. I., D'Odorico, S., Wolfe, A. M., & Dessauges-Zavadsky, M. 2003, ApJ, 582, 596

Li, A., & Draine, B. T. 2001, ApJ, 554, 778

Lidz, A., Faucher-Giguère, C.-A., Dall'Aglio, A., et al. 2010, ApJ, 718, 199

Mackey, J., Bromm, V., & Hernquist, L. 2003, ApJ, 586, 1

Madau P., 1995, ApJ, 441, 18

Madau P., Pozzetti L., Dickinson M., 1998, ApJ, 498, 106

Makovoz, D., & Marleau, F. R. 2005, PASP, 117, 1113

Maller, A. H., Prochaska, J. X., Somerville, R. S., & Primack, J. R. 2001, MNRAS, 326, 1475

Maller, A. H., Prochaska, J. X., Somerville, R. S., & Primack, J. R. 2003, MNRAS, 343, 268

Mannucci, F., Cresci, G., Maiolino, R., et al. 2009, MNRAS, 398, 1915

Matteucci, F., & Francois, P. 1989, MNRAS, 239, 885

Matsuda Y., Yamada T., Hayashino T. et al., 2004, AJ, 128, 569

Matsuda, Y., et al. 2010, MNRAS, L171

Matsuda, Y., Yamada, T., Hayashino, T., et al. 2011, MNRAS, 410, L13

McKee, C. F., & Ostriker, E. C. 2007, ARA&A, 45, 565

McKee, C. F., & Krumholz, M. R. 2010, ApJ, 709, 308

McQuinn, M., Oh, S. P., & Faucher-Giguere, C.-A. 2011, arXiv:1101.1964

Meurer, G. R., Heckman, T. M., Leitherer, C., Kinney, A., Robert, C., & Garnett, D. R. 1995, AJ, 110, 2665

Meyer M. J., et al., 2004, MNRAS, 350, 1195

Miller, G. E., & Scalo, J. M. 1979, ApJS, 41, 513

Mo, H. J., Mao, S., & White, S. D. M. 1998, MNRAS, 295, 319

Moll, S. L., de Grijs, R., Mengel, S., Smith, L. J., & Crowther, P. A. 2009, Ap&SS, 324, 177

Møller P., Warren S. J., 1993, A&A, 270, 43

Møller P., Warren S. J., 1998, MNRAS, 299, 661

Møller P., Warren S. J., Fall S. M. et al., 2002, ApJ, 574, 51

Møller P., Fynbo J. P. U., Fall S. M., 2004, A&A, 422, L33

Monier, E. M., Turnshek, D. A., & Rao, S. 2009, MNRAS, 397, 943

Morton, D. C. 2003, ApJS, 149, 205

Murante, G., Monaco, P., Giovalli, M., Borgani, S., & Diaferio, A. 2010, MNRAS, 405, 1491

Nagamine, K., Wolfe, A. M., Hernquist, L., & Springel, V. 2007, ApJ, 660, 945

Nagamine, K., Choi, J.-H., & Yajima, H. 2010, ApJ, 725, L219

Narayanan, D., Hayward, C. C., Cox, T. J., Hernquist, L., Jonsson, P., Younger, J. D., & Groves, B. 2010, MNRAS, 401, 1613

Narayanan, D., Krumholz, M. R., Ostriker, E. C., & Hernquist, L. 2012, MNRAS, 421, 3127

Nilsson, K. K., Tapken, C., Møller, P., et al. 2009, A&A, 498, 13

Noeske, K. G., Guseva, N. G., Fricke, K. J., Izotov, Y. I., Papaderos, P., & Thuan, T. X. 2000, A&A, 361, 33

Noterdaeme, P., Petitjean, P., Ledoux, C., & Srianand, R. 2009, A&A, 505, 1087

Obreschkow D., Rawlings S., 2009, ApJ, 696, L129

O'Connell, R. W., Gallagher, J. S., III, & Hunter, D. A. 1994, ApJ, 433, 65

Ocvirk, P., Pichon, C., & Teyssier, R. 2008, MNRAS, 390, 1326

Oke J. B., Cohen J. G., Carr M. et al., 1995, PASP, 107, 375

O'Meara, J. M., Tytler, D., Kirkman, D., et al. 2001, ApJ, 552, 718

O'Meara J. M., Chen H.-W., Kaplan D. L., 2006, ApJ, 642, L9

O'Meara, J. M., Burles, S., Prochaska, J. X., et al. 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

Oppenheimer, B. D., Davé, R., Katz, N., Kollmeier, J. A., & Weinberg, D. H. 2012, MNRAS, 420, 829

Origlia, L., Goldader, J. D., Leitherer, C., Schaerer, D., & Oliva, E. 1999, ApJ, 514, 96

Osterbrock, D. E. 1989, Research supported by the University of California, John Simon Guggenheim Memorial Foundation, University of Minnesota, et al. Mill Valley, CA, University Science Books, 1989, 422 p.,

Ostriker, E. C., McKee, C. F., & Leroy, A. K. 2010, ApJ, 721, 975

Papovich C., Dickinson M., Giavalisco M. et al., 2005, ApJ, 631, 101

Paturel, G., Petit, C., Prugniel, P., Theureau, G., Rousseau, J., Brouty, M., Dubois, P., & Cambrésy, L. 2003, A&A, 412, 45

Paturel, G., Theureau, G., Bottinelli, L., Gouguenheim, L., Coudreau-Durand, N., Hallet, N., & Petit, C. 2003, A&A, 412, 57

Pelupessy, F. I., Papadopoulos, P. P., & van der Werf, P. 2006, ApJ, 645, 1024

Penprase, B. E., Prochaska, J. X., Sargent, W. L. W., Toro-Martinez, I., & Beeler, D. J. 2010, ApJ, 721, 1

Péroux, C., Bouché, N., Kulkarni, V. P., York, D. G., & Vladilo, G. 2012, MNRAS, 419, 3060

Petitjean, P., Ledoux, C., Noterdaeme, P., & Srianand, R. 2006, A&A, 456, L9

Petkova, M., & Springel, V. 2009, MNRAS, 396, 1383

Pettini, M., & Bowen, D. V. 2001, ApJ, 560, 41

Pettini, M., Zych, B. J., Steidel, C. C., & Chaffee, F. H. 2008, MNRAS, 385, 2011

Pettini, M., Zych, B. J., Murphy, M. T., Lewis, A., & Steidel, C. C. 2008, MNRAS, 391, 1499

Pontzen, A., et al. 2008, MNRAS, 390, 1349

Pontzen A., Deason A., Governato F. et al., 2009, MNRAS, pp 1943

Powell, L. C., Slyz, A., & Devriendt, J. 2010, arXiv:1012.2839

Prochaska, J. X. 1999, ApJ, 511, L71

Prochaska, J. X., & Burles, S. M. 1999, AJ, 117, 1957

Prochaska, J. X., & Hennawi, J. F. 2009, ApJ, 690, 1558

Prochaska J. X., Wolfe A. M., 1997, ApJ, 487, 73

Prochaska, J. X., & Wolfe, A. M. 2009, ApJ, 696, 1543

Prochaska, J. X., Gawiser, E., Wolfe, A. M., Castro, S., & Djorgovski, S. G. 2003, ApJ, 595, L9

Prochaska, J. X., Herbert-Fort, S., & Wolfe, A. M. 2005, ApJ, 635, 123

Prochaska, J. X., O'Meara, J. M., Herbert-Fort, S., et al. 2006, ApJ, 648, L97

Prochaska, J. X., Chen, H.-W., Dessauges-Zavadsky, M., & Bloom, J. S. 2007, ApJ, 666, 267

Prochaska, J. X., Wolfe, A. M., Howk, J. C., et al. 2007, ApJS, 171, 29

Prochaska, J. X., Worseck, G., & O'Meara, J. M. 2009, ApJ, 705, L113

Prochaska, J. X., Sheffer, Y., Perley, D. A., et al. 2009, ApJ, 691, L27

Prochaska, J. X., O'Meara, J. M., & Worseck, G. 2010, ApJ, 718, 392

Prochter, G. E., Prochaska, J. X., O'Meara, J. M., Burles, S., & Bernstein, R. A. 2010, ApJ, 708, 1221

Pustilnik, S. A., Brinks, E., Thuan, T. X., Lipovetsky, V. A., & Izotov, Y. I. 2001, AJ, 121, 1413

Rafelski M., Wolfe A. M., Cooke J. et al., 2009, ApJ, 703, 2033

Rafelski, M., Wolfe, A. M., & Chen, H.-W. 2011, ApJ, 736, 48

Rafelski, M., Wolfe, A. M., Prochaska, J. X., Neeleman, M., & Mendez, A. J. 2012, arXiv:1205.5047

Rakic, O., Schaye, J., Steidel, C. C., & Rudie, G. C. 2011, arXiv:1109.4944

Rao S. M., Turnshek D. A., Briggs F. H., 1995, ApJ, 449, 488

Rao S. M., Turnshek D. A., Nestor D. B., 2006, ApJ, 636, 610

Rauch, M., Miralda-Escude, J., Sargent, W. L. W., et al. 1997, ApJ, 489, 7

Rauch M., Haehnelt M., Bunker A. et al., 2008, ApJ, 681, 856

Razoumov, A. O. 2009, ApJ, 707, 738

Razoumov, A. O., Norman, M. L., Prochaska, J. X., & Wolfe, A. M. 2006, ApJ, 645, 55

Reddy N. A., Steidel C. C., 2004, ApJ, 603, L13

Reddy N. A., Steidel C. C., 2009, ApJ, 692, 778

Reddy, N. A., Erb, D. K., Pettini, M., Steidel, C. C., & Shapley, A. E. 2010, ApJ, 712, 1070

Rees, M. J., & Ostriker, J. P. 1977, MNRAS, 179, 541

Reimers, D., & Vogel, S. 1993, A&A, 276, L13

Reines, A. E., Johnson, K. E., & Hunt, L. K. 2008, AJ, 136, 1415

Rengarajan, T. N., & Iyengar, K. V. K. 1992, MNRAS, 259, 559

Ribaudo, J., Lehner, N., & Howk, J. C. 2011, arXiv:1105.0659

Robertson, B. E., & Kravtsov, A. V. 2008, ApJ, 680, 1083

Romano, D., Tosi, M., Chiappini, C., & Matteucci, F. 2006, MNRAS, 369, 295

Rudie, G. C., Steidel, C. C., Trainor, R. F., et al. 2012, ApJ, 750, 67

Rutledge R. E., Brunner R. J., Prince T. A. et al., 2000, ApJS, 131, 335

Sage, L. J., Salzer, J. J., Loose, H.-H., & Henkel, C. 1992, A&A, 265, 19

Salim S., Rich R. M., Charlot, S. et al., 2007, ApJS, 173, 267

Sargent, W. L. W., Steidel, C. C., & Boksenberg, A. 1990, ApJ, 351, 364

Savage, B. D., & Sembach, K. R. 1991, ApJ, 379, 245

Scannapieco, C., Wadepuhl, M., Parry, O. H., et al. 2012, MNRAS, 2970

Scannapieco, E., Madau, P., Woosley, S., Heger, A., & Ferrara, A. 2005, ApJ, 633, 1031

Schaye J., 2001, ApJ, 559, L1

Schaye, J. 2006, ApJ, 643, 59

Schaye, J., Aguirre, A., Kim, T.-S., et al. 2003, ApJ, 596, 768

Schaye, J., Carswell, R. F., & Kim, T.-S. 2007, MNRAS, 379, 1169

Schaye, J., Dalla Vecchia, C., Booth, C. M., et al. 2010, MNRAS, 402, 1536

Schlegel D. J., Finkbeiner D. P., Davis M., 1998, ApJ, 500, 525

Schmidt, M. 1959, ApJ, 129, 243

Schmidt, M. 1963, ApJ, 137, 758

Schmitt, H. R., Calzetti, D., Armus, L., Giavalisco, M., Heckman, T. M., Kennicutt, R. C., Jr., Leitherer, C., & Meurer, G. R. 2006, ApJS, 164, 52

Schneider, R., Ferrara, A., Natarajan, P., & Omukai, K. 2002, ApJ, 571, 30

Schruba, A., Leroy, A. K., Walter, F., Sandstrom, K., & Rosolowsky, E. 2010, ApJ, 722, 1699

Schruba, A., Leroy, A. K., Walter, F., et al. 2011, AJ, 142, 37

Shen, S., Wadsley, J., & Stinson, G. 2010, MNRAS, 407, 1581

Silk, J. 1977, ApJ, 211, 638

Simcoe, R. A., Sargent, W. L. W., & Rauch, M. 2004, ApJ, 606, 92

Simcoe, R. A., Sargent, W. L. W., Rauch, M., & Becker, G. 2006, ApJ, 637, 648

Skrutskie, M. F., et al. 2006, AJ, 131, 1163

Smith, B. J., & Hancock, M. 2009, AJ, 138, 130

Solanes, J. M., Giovanelli, R., & Haynes, M. P. 1996, ApJ, 461, 609

Solomon, P. M., Rivolo, A. R., Barrett, J., & Yahil, A. 1987, ApJ, 319, 730

Somerville, R. S., & Primack, J. R. 1999, MNRAS, 310, 1087

Spitzer, L. 1978, New York Wiley-Interscience, 1978. 333 p.,

Springel, V., & Hernquist, L. 2003, MNRAS, 339, 289

Springel, V., White, S. D. M., Jenkins, A., et al. 2005, Nature, 435, 629

Springob, C. M., Haynes, M. P., Giovanelli, R., & Kent, B. R. 2005, ApJS, 160, 149

Stark, A. A., Knapp, G. R., Bally, J., Wilson, R. W., Penzias, A. A., & Rowe, H. E. 1986, ApJ, 310, 660

Steidel, C. C. 1990, ApJS, 74, 37

Steidel, C. C., Adelberger, K. L., Shapley, A. E., et al. 2000, ApJ, 532, 170

Steidel, C. C., Erb, D. K., Shapley, A. E., Pettini, M., Reddy, N., Bogosavljević, M., Rudie, G. C., & Rakic, O. 2010, ApJ, 717, 289

Steidel, C. C., Bogosavljević, M., Shapley, A. E., et al. 2011, ApJ, 736, 160

Steigman, G. 2007, Annual Review of Nuclear and Particle Science, 57, 463

Stewart, K. R., Kaufmann, T., Bullock, J. S., Barton, E. J., Maller, A. H., Diemand, J., & Wadsley, J. 2010, arXiv:1012.2128

Strong, A. W., & Mattox, J. R. 1996, A&A, 308, L21

Storrie-Lombardi L. J., Wolfe A. M., 2000, ApJ, 543, 552

Suda, T., Yamada, S., Katsuta, Y., et al. 2011, MNRAS, 412, 843

Sutherland W., Saunders W., 1992, MNRAS, 259, 413

Tacconi, L. J., et al. 2010, Nature, 463, 781

Taylor, C. L., Kobulnicky, H. A., & Skillman, E. D. 1998, AJ, 116, 2746

Tescari, E., Viel, M., Tornatore, L., & Borgani, S. 2009, MNRAS, 397, 411

Thuan, T. X., & Izotov, Y. I. 1997, ApJ, 489, 623

Thuan, T. X., & Izotov, Y. I. 2005, ApJ, 627, 739

Thuan, T. X., Hibbard, J. E., & Lévrier, F. 2004, AJ, 128, 617

Thuan, T. X., Lecavelier des Etangs, A., & Izotov, Y. I. 2005, ApJ, 621, 269

Tielens, A. G. G. M. 2005, The Physics and Chemistry of the Interstellar Medium, by A. G. G. M. Tielens, pp. . ISBN 0521826349. Cambridge, UK: Cambridge University Press, 2005.,

Tolstoy, E., Saha, A., Hoessel, J. G., & McQuade, K. 1995, AJ, 110, 1640

Toomre, A. 1964, ApJ, 139, 1217

Tully, R. B. 1994, VizieR Online Data Catalog, 7145, 0

Tumlinson, J., Shull, J. M., Rachford, B. L., et al. 2002, ApJ, 566, 857

Tumlinson, J., Thom, C., Werk, J. K., et al. 2011, Science, 334, 948

van den Bergh, S. 1976, ApJ, 206, 883

van de Voort, F., Schaye, J., Altay, G., & Theuns, T. 2012, MNRAS, 421, 2809

van Dishoeck, E. F., & Black, J. H. 1986, ApJS, 62, 109

van Zee, L., & Haynes, M. P. 2006, ApJ, 636, 214

van Zee, L., Westpfahl, D., Haynes, M. P., & Salzer, J. J. 1998, AJ, 115, 1000

van Zee, L., Skillman, E. D., & Salzer, J. J. 1998, AJ, 116, 1186

Vanzi, L. 2003, A&A, 408, 523

Vanzi, L., & Sauvage, M. 2004, A&A, 415, 509

Vanzi, L., Hunt, L. K., Thuan, T. X., & Izotov, Y. I. 2000, A&A, 363, 493

Vanzi, L., Hunt, L. K., & Thuan, T. X. 2002, A&A, 390, 481

Vanzi, L., Cresci, G., Telles, E., & Melnick, J. 2008, A&A, 486, 393

Verheijen M. A. W., 2001, ApJ, 563, 694

Vogt, S. S., Allen, S. L., Bigelow, B. C., et al. 1994, Proc. SPIE, 2198, 362

Vollmer, B., Braine, J., Pappalardo, C., & Hily-Blant, P. 2008, A&A, 491, 455

Wadsley, J. W., Veeravalli, G., & Couchman, H. M. P. 2008, MNRAS, 387, 427

Wall, W. F. 2007, MNRAS, 379, 674

Walter F., Brinks E., de Blok W. J. G. et al., 2008, AJ, 136, 2563

Warren S. J., Møller P., Fall S. M. et al., 2001, MNRAS, 326, 759

Weatherley S. J., Warren S. J., Møller P. et al., 2005, MNRAS, 358, 985

Whitmore, B. C., et al. 2005, AJ, 130, 2104

Wilson, C. D., et al. 2008, arXiv:0812.1718

Windhorst, R. A., Cohen, S. H., Hathi, N. P., et al. 2011, ApJS, 193, 27

Wise, J. H., & Abel, T. 2008, ApJ, 685, 40

Wise, J. H., & Abel, T. 2010, arXiv:1012.2865

Wise, J. H., Turk, M. J., Norman, M. L., & Abel, T. 2012, ApJ, 745, 50

Wolfe A. M., Chen H.-W., 2006, ApJ, 652, 981

Wolfe A. M., Turnshek D. A., Smith H. E. et al., 1986, ApJS, 61, 249

Wolfe A. M., Prochaska J. X., Gawiser E., 2003, ApJ, 593, 215

Wolfe A. M., Gawiser E., Prochaska J. X., 2005, ARA&A, 43, 861

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

Wolfire, M. G., Hollenbach, D., & McKee, C. F. 2010, ApJ, 716, 1191

Wong, T., & Blitz, L. 2002, ApJ, 569, 157

Wood, K., Haffner, L. M., Reynolds, R. J., Mathis, J. S., & Madsen, G. 2005, ApJ, 633, 295

Woosley, S. E., & Weaver, T. A. 1995, ApJS, 101, 181

Yoshida, N., Bromm, V., & Hernquist, L. 2004, ApJ, 605, 579

Young, J. S., Xie, S., Kenney, J. D. P., & Rice, W. L. 1989, ApJS, 70, 699

Zaghloul, M. R. 2007, MNRAS, 375, 1043

Zheng, Z., & Miralda-Escudé, J. 2002, ApJ, 568, L71

Zwaan, M. A., & Prochaska, J. X. 2006, ApJ, 643, 675

Zwaan M. A., van der Hulst J. M., Briggs F. H. et al., 2005a, MNRAS, 364, 1467

Zwaan M. A., Meyer M. J., Staveley-Smith L. et al., 2005b, MNRAS, 359, L30