Università degli Studi di Trieste

PhD Thesis

The denser absorbers in the IGM and HI in galaxies

Settore scientifico-disciplinare:
Astronomia e Astrofisica

Ph.D. program Coordinator:
Prof. Francesco Longo

PhD student:
Serafina Di Gioia

Supervisor:
Prof. Stefano Cristiani
Co-supervisor:
Dr. Gabriella De Lucia

XXXIII ciclo

Short summary

My PhD thesis focuses on the study of the connection between the densest absorbers in the Intergalactic Medium (IGM), in particular the Damped Lyman alpha systems (DLAs), and the atomic hydrogen (HI) in galaxies. As tools for my research I used different versions of the 'state-of-the-art' semi-analytic model Galaxy Evolution and Assembly (GAEA), coupled to cosmological N-body simulations.

In the first part of this thesis I analyze the predictions of the GAEA model for the DLA observables in the redshift range $2 < z < 3$. I studied, in particular, the BR run presented in Xie et al. (2017), that adopts a partitioning scheme based on the Blitz-Rosolowsky relation, derived from observations of galaxies in the local Universe (Blitz & Rosolowsky, 2006). I have chosen to investigate the redshift range $2 < z < 3$ for two reasons: first, this epoch is very important for galaxy evolution (at $z \sim 2$ we have the peak of the SFR density) and second in this redshift range we have the most complete DLA surveys and the most precise DLA measurements. I estimated the DLA column density distribution function (CDDF), the comoving HI density associated to DLAs ($\Omega_{\rm HI}^{DLA}$), their typical host halo mass, the average impact parameter and the DLA metallicity distribution. At $z < 3$ our fiducial model predicts the correct shape of the CDDF, but its normalization falls short of the observations, with the discrepancy increasing at higher redshift. The agreement of the model with observations is significantly improved increasing both the HI masses and the disk radii of model galaxies by a factor 2. Haloes with $M_{200} \geq 10^{11} M_{\odot}$ give the major contribution to $\Omega_{\rm HI}^{DLA}$, and the typical DLA host halo mass is $\sim 10^{11} M_{\odot}$. The simulated DLA metallicity distribution is in relatively good agreement with observations, but our model predicts an excess of DLAs at low metallicities. The results of this work suggest possible improvements for the adopted modelling of the angular momentum of galactic disks, of the filtering mass and metal ejection in low-mass haloes.

In the second part of this thesis I focus on improving the modelling of HI in the GAEA model, to get a better agreement with DLA observations at $z < 3$. In particular I work on a new version of the model (Xie et al., 2020b), that includes a new treatment for the ram pressure stripping and implements a different modelling tracing the specific angular momentum of the gaseous disk in galaxies. We refer to this model version as RPS model. As

a starting point I modify the assumed reionization history from an early to a late-reionization scenario. Then I test different models for the filtering mass M_F (e.g. Macciò et al., 2010; Kim et al., 2015). The change of the reionization scenario does not affect in a significant way the galaxy properties at the redshift of interest. The effect of the modified prescription for the filtering mass is larger and the alternative prescriptions considered bring very different results in terms of HI density evolution. Our new reference prescription for M_F produces a better distribution of the HI in halos of different masses, increasing the relative contribution of intermediate/low mass haloes to the HI density, and consequently it leads to a shape of the CDDF that is more in agreement with observations. However, this new implementation does not solve the problem of the different normalization of the simulated CDDF with respect to the observed one at $z = 2$. All the modifications considered change only up to 10% the HI content in the model galaxies, in the redshift range of interest. Finally, I investigate the effects of adopting different prescriptions for the partitioning of cold gas into molecular and atomic. In particular I adopt four cold gas partitioning schemes: the first one is based on the BR law (Blitz & Rosolowsky, 2006), the second one is based on the results by Gnedin & Kravtsov (2011),the third one is based on the analytical model proposed by Krumholz et al. (2008, 2009a), that relates the HI content of a galaxy to its metallicity, and the last one is a modified version of the BR prescription. The GAEA-RPS runs based on the GK and KMT predict very different galaxy properties with respect to the run based on the BR original/modified prescription. The **GAEA-RPS-mod-BR-evo**P_0 model is defined as the new reference model, since its predictions for DLA observables and HI in galaxies are the ones more in agreement with the observations. For this model I test also: i) the predictions regarding the molecular gas content associated to galaxies and DLAs, ii) how the DLAs statistics change considering not only the cold gas in the ISM but also the warm/cold gas in the ejected component associated to galaxies.

Contents

Chapter 1

Introduction

The subject of my PhD Thesis is the study of the nature and origin of the densest absorbers in the Intergalactic Medium (IGM). In particular I studied the connection between the densest absorbers in the IGM, the Damped Lyman-α systems(DLAs), and galaxies, by means of a 'state of the art' semi-analytic model (GAEA). This is an important step in order to improve the modelling of important processes for galaxy formation and evolution thanks to the constraints coming from DLA observations.
The IGM nowadays shows itself as a 'pervasive medium of ionized gas', mostly made of intergalactic hydrogen. The sensitivity of the current IGM observations and the large observational statistics make the IGM a unique tool for both cosmology and galaxy evolution studies. Recent studies show that $\sim 80\%$ of the baryonic matter in the high-redshift Universe is in the IGM (Péroux & Howk, 2020; Crighton et al., 2015).

In the next sections I will summarize the cosmological background.

1.1 Cosmological context

The effort of the scientific community, from astronomers to theoretical physicists, has lead to the formulation of the Standard Cosmological Model, that is able to explain an increasingly large set of observations, from the clustering of galaxies on large scales to the temperature anisotropies, measured by the CMB and the Lyman Alpha forest statistics (McCrea, 1972; Blumenthal et al., 1984; Cole et al., 2005; Hinshaw et al., 2013; Slosar et al., 2013; Planck

Collaboration et al., 2018).

Nowadays even outside the astrophysical community people know that we live in an expanding and accelerating Universe. The first test of the Universe expansion is the empirical Hubble law, according to which distant galaxies are receding from us at a velocity proportional to their distance from the Earth:

$$v_r = H_0 D \tag{1.1}$$

The slope of the relation between the distance (D) and recession velocity (v_r) is defined to be the Hubble constant, H_0 (the most recent estimate is $H_0 = (67.4 \pm 0.5) kms^{-1}Mpc^{-1}$).

Evidence supporting cosmic acceleration has grown since its discovery in 1998, by two independent projects, the Supernova Cosmology Project and the High-Z Supernova Search Team, both based on the analysis of data from type Ia Supernovae (SNe) (Riess et al., 1998; Garnavich et al., 1998; Perlmutter et al., 1999). A large number of independent observables support now this paradigm. Example of different observables, alternative to SNe, that constrain the cosmic acceleration, are: the detection of Baryonic Acoustic Oscillations in the Lyman forest (BAO, Aubourg et al., 2015), weak lensing (Mandelbaum, 2018), galaxy cluster counts (Allen et al., 2003), and redshift space distortion measurements from galaxy surveys (Guzzo et al., 2008).

The dynamics and geometry of the observed Universe has proven to be well described by the Einstein's General Relativity Theory. In this theory all dynamical information is included in the metric, $g_{\mu\nu}$, and in the energy tensor (the space-time structure is strictly linked to the distribution of matter in space). Einstein's theory together with the assumption that the Universe is isotropic and homogeneous on large scale (the Cosmological Principle), as probed by the measurements of the Cosmic Microwave Background (CMB) temperature anisotropies, represents the basis of the standard cosmological model. By assuming the cosmological principle, we assume the existence of a fundamental observer, arbitrary placed, and this implies that he lives on a three-dimensional hyper-surface where all the locally defined properties (density, Universe expansion rate etc.) evolve uniformly according to a universal time, called *cosmic time*.

So we can describe the spacetime associated to the Universe with the metric of the maximally symmetric varieties, the Friedman-Robertson-Walker's

metric (FRW metric):

$$ds^2 = c^2dt^2 - \frac{a^2(t)}{1-kr^2}[dr^2 + r^2d\theta^2 + r^2\sin^2\theta d\phi^2] \tag{1.2}$$

Still, two unknown elements need to be added: namely a non-relativistic species of matter that does not have electromagnetic interactions, the Cold Dark Matter (CDM) and an unknown source of energy (Dark Energy, DE), the *Cosmological Constant* (Λ).The first one is needed to explain the formation and dynamics of cosmic structures while the cosmological constant is necessary to propel the latter stage of accelerated expansion. These ingredients enter in the name of the standard cosmological model: the ΛCDM model.

1.1.1 ΛCDM model

This model represents a particular parametrization in the class of the Big Bang models.

The parametrization of the ΛCDM model is based on 6 independent parameters. Several observational studies are devoted to the measurement of these parameters, at the epoch of CMB (e.g. WMAP, Planck), and in the local Universe (e.g BOSS, DESI). The 6 parameters are: the dark matter density $\Omega_c h^2$, the baryon density $\Omega_b h^2$, the age of the Universe t_0, the reionization optical depth τ, the initial super-horizon amplitude of curvature perturbations A_s at $k = 0.05\mathrm{Mpc}^{-1}$, and the primordial spectral index n_s. The best fit values of these parameters, in the Planck analysis (Planck Collaboration et al., 2018), are found to be $\Omega_b h^2 = 0.0224 \pm 0.0001$, $\Omega_c h^2 = 0.120 \pm 0.001$, $100\theta_\star = 1.0411 \pm 0.0003$, $\tau = 0.054 \pm 0.007$,$n_s = 0.965 \pm 0.004$, $ln(A_s) = 3.094 \pm 0.034$. From these constraints summed to the assumptions at the basis of the ΛCDM model it can be derived also a value for the matter fluctuation amplitude $\sigma_8 = 0.811 \pm 0.006$ and the Hubble parameter $h = 0.674 \pm 0.005$.

The recent estimate of the cosmological parameters by the Planck mission provides us with the total energy density budget of the Universe, where the main component is dark energy, a repulsive force driving the Universe expansion (represented by the constant Λ), accounting for $\sim 68\%$ of the total budget. The rest of the Universe is composed of matter, with baryons accounting only for almost $\sim 5\%$, and Dark Matter (DM) for the remaining part ($\sim 27\%$). The nature of DM is still uncertain, but the most accepted theory states that it is composed of weakly interacting massive particles

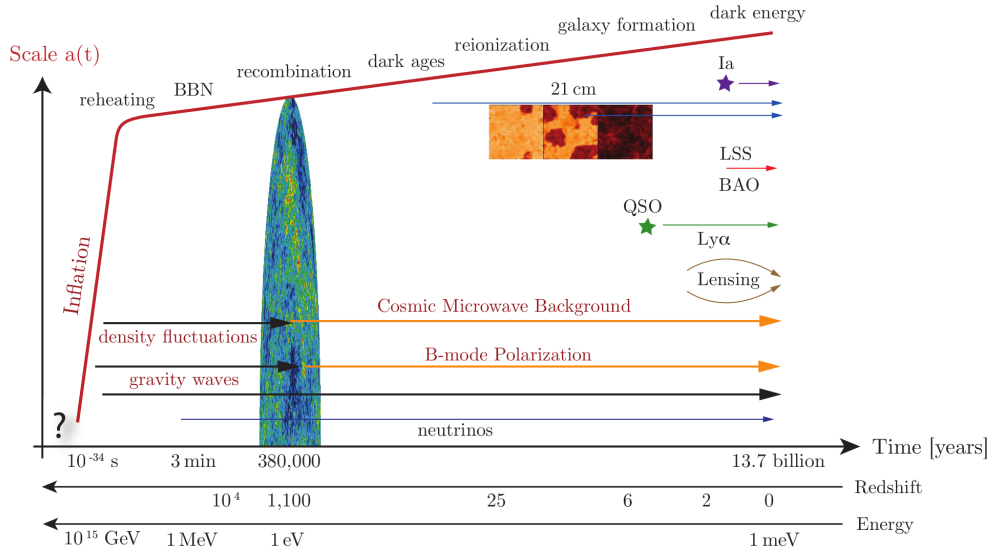

Fig. 1.1: History of the universe. In this schematic the key events in the history of the universe are presented, with their associated time and energy scales. The figure also illustrates several cosmological probes that provide us with information about the structure and evolution of the universe. Acronyms: BBN (Big Bang Nucleosynthesis), LSS (Large-Scale Structure), BAO (Baryon Acoustic Oscillations), QSO (Quasars), Lyα (Lyman-alpha), CMB (Cosmic Microwave Background), Ia (Type Ia supernovae), 21cm (hydrogen 21cm-transition). Credits: D. Baumann

(WIMPs) that do not interact electromagnetically and are the products of particle interactions in the early Universe.

The dynamical evolution of the Universe is described by the Friedmann equations (derived from the Einstein's equations once applied the assumption at the basis of the ΛCDM model) :

$$\dot{a}^2 = \frac{8\pi G}{3}\rho a^2 - kc^2 + \frac{1}{3}\Lambda a^2 \tag{1.3}$$

$$\ddot{a} = -\frac{4\pi G}{3}\left(\rho + \frac{3p}{c^2}\right)a - 2\frac{\Lambda}{3}a \tag{1.4}$$

where a is the scale factor

In Fig 1.1 it is shown the history of the Universe, according to the Standard ΛCDM model.

In the standard cosmological model everything begins with the Big Bang. Georges Lemaître first noted in 1927 that an expanding universe could be

traced back in time to an originating single point, as written in the "Theory of primeval atom". For much of the rest of the $20^{\rm th}$ century the astrophysical community was divided between supporters of the Big Bang model and the rival steady-state model. Since the discovery of a cosmic microwave background by Penzias and Wilson (1965) a wide range of empirical evidences strongly favored the Big Bang, which is now universally accepted.

The ΛCDM model assumes that, during its earliest evolutionary stages, the Universe was extremely hot and dense, the so called 'cosmic plasma'. Initially the hot plasma was made of photons, electrons, positrons and neutrinos, with an average temperature $T > 10^{13}$ K. At this early epoch the standard laws of thermodynamics are not valid and theoretical physicist have proposed different models to describe the physics regulating the evolution of the primordial Universe. It is believed that during this epoch, two important processes took place: baryogenesis, that created an asymmetry in favor of baryons (and against anti-baryons), and inflation.

The theory of inflation was introduced to explain some properties of the observed Universe which are not predicted by the Big Bang model (Barrow & Turner, 1981). This unexpected properties of the Universe have represented for many years the most fascinating open problems in Cosmology.

The first is the horizon problem: the COBE experiment, which observed the CMB anisotropies in 1991, showed that the Universe temperature at the epoch of the last scattering surface is homogeneous, with fluctuations of the order 10^{-4} (Mather et al., 1990; Smoot, 1999), even if at this epoch the Universe was formed by causally disconnected regions. This can be naturally explained if the Universe originated from a very small region where all matter and energy were in close and uniform contact and then suddenly expanded in a very fast way.

The second problem is the flatness problem: since we observe that the Universe has a mean density close to the critical density today, after $\sim$ 14 billion years of expansion and evolution, it must have been even closer at earlier times. For instance, this requires the density at the Planck time (within 10^{-43} seconds of the Big Bang) to be within 1 part in 10^{57} of the critical density. Explaining this without inflation requires a fine tuned set of initial values for the matter and energy density, and small deviations from these values would have had extreme effects on the currently observed Universe.

The last problem is the monopole problem: if the early Universe was very

hot as assumed, a large number of heavy magnetic monopoles would have been produced, and should have survived to present day. Observationally, we do not see any evidence of their existence. An inflationary period, happened just after their formation, would have diluted them significantly.

Around 10^{-32} seconds after the Big Bang the inflation begins and leads to an exponential growth of the Universe.

From 10^{-32} to 10^{-12} seconds after the Big Bang the inflationary expansion forced matter to cool and pass through a number of symmetry-breaking phase transitions (SUSY breaking, hadrosynthesis). At the end of the inflation the quantum fluctuations present in the cosmic plasma gave origin to density fluctuations.

Around 10^{-10} seconds after the Big Bang the EW symmetry broke and the bosons Z ans $W^{\pm}$ acquired masses.

From 10^{-10} seconds after the Big Bang to today the history of the universe is based on well understood and experimentally tested laws of particle physics, nuclear and atomic physics and gravity.

After the EW phase transition the neutrino decoupled from matter when the Universe reached an average temperature of 1 MeV.

Shortly after at 1 second, the temperature drops below the electron rest mass and electrons and positrons annihilate efficiently leaving only the matter generated by the baryon/anti-baryon asymmetry.

Around 200s after the Big Bang, at a temperature of 0.1 MeV the strong interactions became effective and protons and neutrons combined into the light elements (H, He, Li) during Big Bang Nucleosynthesis(BBN).

Atoms formed at $z \sim 1100$ when the Universe reached a low enough temperature for the cosmic plasma to recombine. Actually, due to the large value of the cosmic photon-to-baryon ratio, $\eta = 10^9$, this process was delayed until the temperature dropped to 0.29 eV, about 50 times lower than the binding energy of hydrogen. Cosmic recombination proceeded far out of equilibrium because of a "bottleneck" at the n= 2 level of hydrogen since atoms can only reach the ground state via slow processes including the two-photon decay or Lyman-alpha resonance escape. As the recombination rate rapidly became smaller than the expansion rate, the reaction could not reach completion and a small relic abundance of free electrons was left behind at $z < 1000$.

When recombination starts photons decoupled and formed the free-streaming

cosmic microwave background. The CMB filled the universe with a red, uniformly bright glow of blackbody radiation, but later the temperature dropped and the CMB shifted to the infrared. At that epoch, the universe would appear as a completely dark place to human eyes. A long period of time has to pass until the first cosmic structures collapsed, forming the first stars that filled the universe with the first light ever emitted, that was not part of the CMB, ending the epoch known as the *Dark Ages* of the universe.

The mechanism that guides the evolution of Cosmic structures is *gravitational instability*. Stated simply the key idea is that: *If the matter in the Universe is initially distributed irregularly the overdense regions attract matter towards them and become overdense and then can collapse or not* .

Structure formation occurs when a density perturbation reaches a critical density, and decouples from the expansion of the Universe, undergoing a collapse due to its own gravity. The collapse process can be described with several theoretical models. The simplest analytical solution is obtained assuming a spherical symmetry. This solution is calculated assuming that the collapsing spherical overdensity can be treated as a closed Universe, using the Friedmann equations. This leads to an analytic description of the radius and the time describing the collapse of the spherical overdensity, and to the result that the perturbation can be considered collapsed when the initial overdensity is ~ 180 times the average density of the Universe. Actually the collapse of the DM perturbations involves additional complexities, such as the non spherical density distribution,the shell crossing, or rotational torques in the collapsing overdensity due to tidal fields, which require a numerical treatment. Numerical simulations confirm that, on average, structures collapse and separate from the cosmological background when their density is around ~ 100 times the mean density of the Universe. After the first shell crossing the virialization process begins. During virialization each shell in the spherical distribution, oscillating, interacts with the others, exchanging energy. Through these processes the system relaxes towards virial equilibrium, becoming “virialized”. We define a virialized DM structure a “DM halo”. Numerical simulations show that on average the structure of the virialized DM halos is quite well described by the Navarro – Frenk – White (NFW) density profile (Navarro et al., 1995, 1996; Ludlow et al., 2013).

After that the halo becomes virialized it can grow further through accretion of material from the external regions, or through mergers with other

halos. When a small halo is merged or accreted into a larger one, without being destroyed by its tidal field, it becomes a self-bound substructure, and we call it a subhalo.

While the growth of DM perturbations is prevented only by the Universe expansion, that tends to dilute the over-densities, the growth of the first galaxies is prevented by the initial high temperature. The first generation of stars formed in relatively low-mass galaxies. Massive galaxies, and even more massive structures such as galaxy clusters, formed later.

The formation and evolution of galaxies is strictly linked to that of the DM halos.

1.1.2 From the Reionization to nowadays

With the birth of the first galaxies and stars the Dark Ages ended, around $z \sim 30$. The first population of luminous stars and galaxies generated ultraviolet (UV) radiation through nuclear reactions. They were not the only sources of UV photons at that time: an early population of accreting black holes (quasars) and the decay/annihilation of dark matter particles also generated some amount of UV light. The UV photons with $E_\gamma > 13.6eV$ then were able to ionize hydrogen atoms in the medium surrounding the UV sources. This process is known as "cosmic reionization" and it represents the second major change in the ionization state of hydrogen (after recombination).

Three epochs of reionization (EoR) of the IGM may be identified, one for hydrogen and two for helium [1]. There is strong evidence full Helium reionization occurred at $z \sim 3.5$ (Planck Collaboration et al., 2018), but the time of Hydrogen reionization is still debated.

In the simplest picture of the Hydrogen reionization each source first produced an ionized region around itself (this phase is known as *pre-overlap* phase) which overlapped and percolated into the IGM during the *overlap* phase. The process of overlapping completed around $z \sim 7$ when the neutral hydrogen fraction fell to values 10^{-4}. After that, a never-ending *post-overlap* phase started as we can deduce from observing that the Universe is largely ionized at the present epoch. Reionization by UV radiation was also accompanied by heating: electrons released by photo-ionization deposited the photon energy in excess of 13.6 eV into the IGM. During reionization, photoioniza-

[1] The first phase happened around $z \sim 6$ while the second at $z \sim 3$.

tion heating from the rising UVB can increase the gas temperature of about 2×10^4 K in ionized regions. This temperature is equal to the virial temperature of a bound halo of mass $M_{200} \sim 10^{8.5} M_{\odot}$ at the reionization redshift. The equivalence between the external radiation field and the internal binding energy can have two effects: firstly, it can photo-evaporate gas from the small potential wells of the lowest mass galaxies (in the minihalo/low mass end of atomic cooling halo regime); secondly, it can increase the gas pressure, associated with the increase in $T_{\rm IGM}$, increasing the Jeans scale for gas collapse (Gnedin, 2000; Okamoto et al., 2008b; Noh & McQuinn, 2014) and inhibiting gas accretion onto galaxies. Both these effects can combine to suppress further star formation under a critical halo mass value.

From the measurements of the IGM Ly-α optical depth in high-redshift QSOs (e.g. Becker et al., 2001, 2007, 2015; Pentericci et al., 2002; Fan et al., 2006; Djorgovski et al., 2001) and the recent CMB measurements (Planck Collaboration et al., 2018) of intergalactic Thomson scattering a late-reionization scenario is favoured ($z_{reion} = 7.5 - 9$).

Thanks mainly to the analysis of the CMB optical depth performed by Planck Collaboration (2018) and to the high-z QSOs observations it is also clear now that the EoR was a very rapid event. The overlap phase transition lasted for a short redshift interval $\Delta z \leq 2.8$, and it also was a patchy process (Becker et al., 2015).

Chapter 2

InterGalactic Medium

As the name suggests the Intergalactic Medium (IGM) is made of intergalactic gas, which covers a very large dynamical range in scales and densities and is mostly probed in absorption, with few current notable experiments revealing it also in emission (e.g. MUSE, Bacon et al., 2021). The history of the IGM studies is strictly linked to that of quasars (QSOs). Soon after the first identification of a quasar the first absorption lines in a QSO spectrum were observed (in particular in the spectrum of 3C 191, firsty studied by Burbidge et al., 1966). The existence of absorption features in QSO spectra was predicted by Gunn & Peterson (1965) as a sign of the absorption of photons by the intergalactic neutral hydrogen in its ground state. Differently from what expected by Gunn & Peterson (1965) the absorption spectra of the first quasars at $2 < z < 3$ did not reveal a big absorption through in the region blue-ward of the (Ly–α) emission from the QSO, but instead the presence of particular absorption lines, due to metals, as predicted by Bahcall & Salpeter (1965). The absence of the big absorption through, which would be present if the IGM were significantly neutral at the observed redshift, implies that the IGM is fully ionized at $z \sim 3$.

2.1 IGM observables

To understand how the absorption features in quasar spectra are generated we can imagine that neutral hydrogen clouds at different positions along the line of sight, between us and the QSO, absorb the photons at different

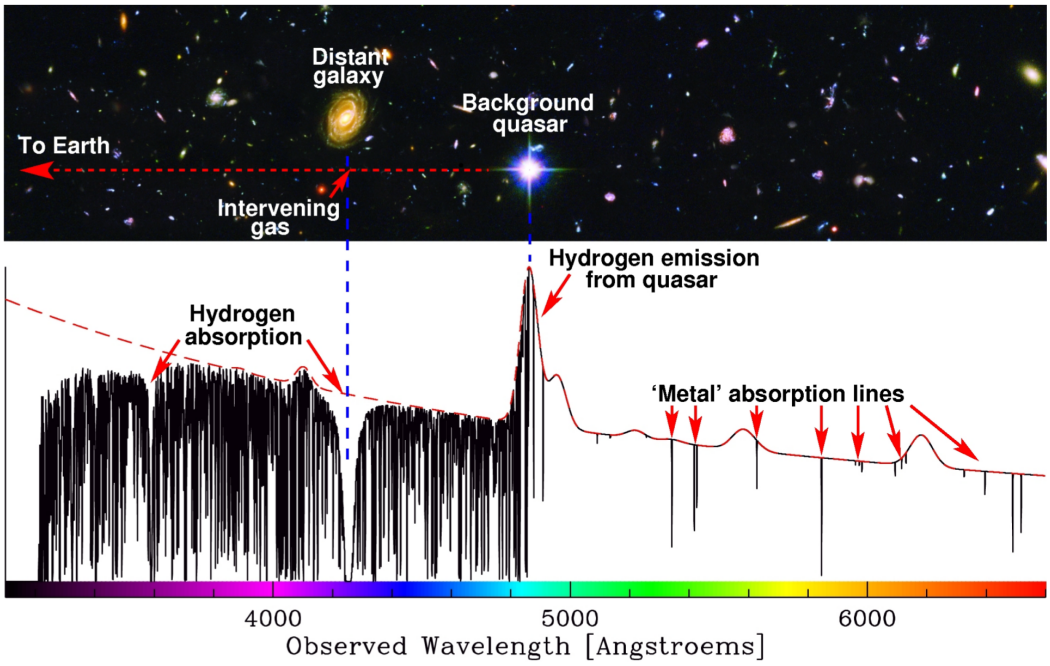

Fig. 2.1: Synthetic spectrum of a QSO, showing an absorption feature corresponding to intervening gas on the l.o.s. The intrinsic continuum is plotted in red. We can see the Ly-α forest blueward of the quasar peak emission, and the metal absorption lines redward. Credits: J. Webb

wavelengths (due to their different redshift). Then each individual cloud leaves its fingerprint as an absorption line at a different position in the observed spectrum blue-ward of the Ly–α emission line of the quasar, as illustrated in fig. 2.1

But how do we model the physics of the absorption of photons by a medium? This is the subject of the Radiative Transfer studies. Assuming that a beam of photons of intensity $I(\nu)$ passes through an absorbing medium, of thickness ds, we will observe an attenuation of the intensity equal to:

$$dI(\nu) = n(s)\,\sigma(\nu)\,I(\nu)\,ds \tag{2.1}$$

where $n(s)$ is the number density of absorbing atoms and $\sigma(\nu)$ is the cross-section of the absorption, which depends on the frequency (ν).

We can rewrite the cross-section of the absorbing atoms transiting between the states $j - k$ as an integrated cross-section over all the frequencies

(σ_{jk}) convolved with the line profile ($\phi(\nu)$):

$$\sigma(\nu) = \sigma_{jk}\phi(\nu) \tag{2.2}$$

In the case of the Lyman-α absorption features in quasar spectra the line profile results from the combination of two different broadening mechanisms:

1. the Doppler broadening

2. the collisional+natural broadening

The Doppler broadening is due to the motion of individual absorbing atoms. If we assume the distribution of velocities to be Maxwellian we can write the line profile as:

$$\phi(\nu) = \frac{1}{\sqrt{\pi}\Delta\nu_D} e^{-\left(\frac{\Delta\nu}{\Delta\nu_D}\right)^2} \tag{2.3}$$

where $\Delta\nu_D = \frac{\nu_0}{c} b_{th}$, ν_0 is the frequency of the line centre, and $b_{th} = \sqrt{\frac{2kT}{m}}$ is the thermal Doppler parameter.

The natural broadening arises because excited states have a finite lifetime. If collisions are frequent enough, they will randomize the phase of the emitted radiation and (effectively) shorten the lifetime further. Assuming a frequency of collision ν_0, then the line profile is given by:

$$\phi(\nu) = \frac{\Gamma/4\pi^2}{(\nu - \nu_0)^2 + \Gamma/4\pi^2} \tag{2.4}$$

The convolution of the two line profiles described above gives the Voigt profile, which can be fitted by a Voigt function with integral:

$$\int_{-\infty}^{\infty} dy \frac{a}{\pi^{3/2}\Delta\nu_D} \frac{e^{-y^2}}{(x-y)^2 + a^2}^1 \tag{2.5}$$

The Voigt profile is fully specified by 3 parameters:

[1]The Voigt integral cannot be evaluated analytically and so it is generated numerically. Humlicek (1979) gives a *Fortran* subroutine for H(a, x), and numerical tables are given by Finn & Mugglestone (1965). In the quasar absorption line literature, the programs VPFIT (Carswell & Webb, 2014) and FITLYMAN (Fontana & Ballester, 1995) are in widespread use for fitting Voigt profiles.

- position in the velocity space (or redshift)
- column density ($N_{\rm HI}$)
- line width, expressed in term of the Doppler parameter (b)

The Lyman-α forest as a whole can then be conveniently characterized in terms of the distribution function of these three quantities.

From the measured flux in the Lyman-α forest region of the quasar spectra we can derive the *opacity* of the IGM at a given λ as:

$$\tau_{\rm HI} \; / \; f_\lambda = f_C e^{-\tau_{\rm HI}(\lambda)}$$

where f_λ is the observed flux of the QSO, and f_C is the unabsorbed flux of the QSO.

If the absorption line is not well resolved [2], instead of fitting the line with the Voigt profile astronomers measure the equivalent width (W) as:

$$W_{obs} = \int d\lambda (1 - e^{-\tau(\lambda)}) \tag{2.6}$$

Before going to describe the statistical and physical properties of the Lyman-α forest, I want to introduce 2 classifications of the absorption features. The first classification is historically motivated by the tentative of distinguish between intrinsic and extrinsic features in QSO spectra. According to this classification scheme, the absorbers were associated to three different classes:

1. the broad absorption lines (BAL), thought to be intrinsic, are linked to gas outflowing from the AGN, by analogy with P-Cyg profiles in stars;
2. the sharp metal absorption lines, typically identified on the basis of MgII, CIV, SiIV and NV doublets, sometimes associated with the QSO for absorption redshifts close to the emission redshift, are in most cases intervening. This class of absorbers displays significant clustering on scales typical of galaxies. Both the metallicity and the clustering suggest a connection with galactic structures;
3. the Lyman$-\alpha$ forest, ascribed to a sort of primordial intergalactic gas.

[2] An absorption line is considered resolved when the entire profile of the line is detected.

The second classification is based on the hydrogen column density, $N_{\rm HI}$. In this classification scheme we have 3 classes of absorbers:

- Lyman-α forest lines - absorption systems with $N_{\rm HI} < 10^{17} atoms\,{\rm cm}^{-2}$
- Lyman Limit Systems (LLS)- absorbers with $10^{17.2} \leq N_{\rm HI} < 10^{20.3} atoms\,{\rm cm}^{-2}$
- Damped Lyman-α systems (DLA) - absorbers with $N_{\rm HI} \geq 10^{20.3} atoms\,{\rm cm}^{-2}$

I am going to introduce now important observables of the IGM that are linked to the statistics of the absorption features.

The simplest one is the number density of lines per unit redshift ($n(z)$) or *line density*.

More complex is to measure the Column Density Distribution Function (CDDF) which expresses the number of absorbers as a function of column density ($N_{\rm HI}$) and redshift, along a given path through the Universe:

$$f(N_{\rm HI}, X) dX dN_{\rm HI} = n_{abs}(N_{\rm HI}, X), \tag{2.7}$$

where the absorbing path dX is defined as $dX = \frac{H_0}{H(z)}(1+z)^2 dz$, in terms of the redshift path dz.

Constraining the CDDF at different redshifts has required significant observational effort, mainly because of the very large dynamic range of the measurements ($10^{12} < N_{\rm HI} < 10^{22}{\rm cm}^{-2}$) which have necessitated surveys optimized for particular ranges in column density. This requirement is easily understood by considering the CURVE of GROWTH (CoG) [3] of the Lyman-α line of neutral hydrogen, showed in Fig 2.2.

Looking at the CoG we can identify three regimes:

- optically thin lines or 'linear regime'
- lines saturated in the core or 'flat regime'(for a Voigt profile, this part of the profile is determined by the Doppler broadening)
- lines saturated out to frequencies/wavelengths where the line profile is determined by the Lorentzian wings

[3]The equivalent width reduces the information in the line to one number, proportional to the area of the line, or its strength; the line profile information is lost. However, if you measure lines from the same ion with different f-values, then you can construct a curve of growth and deduce the column density, N, and temperature, T , of the ions in the absorbing cloud.

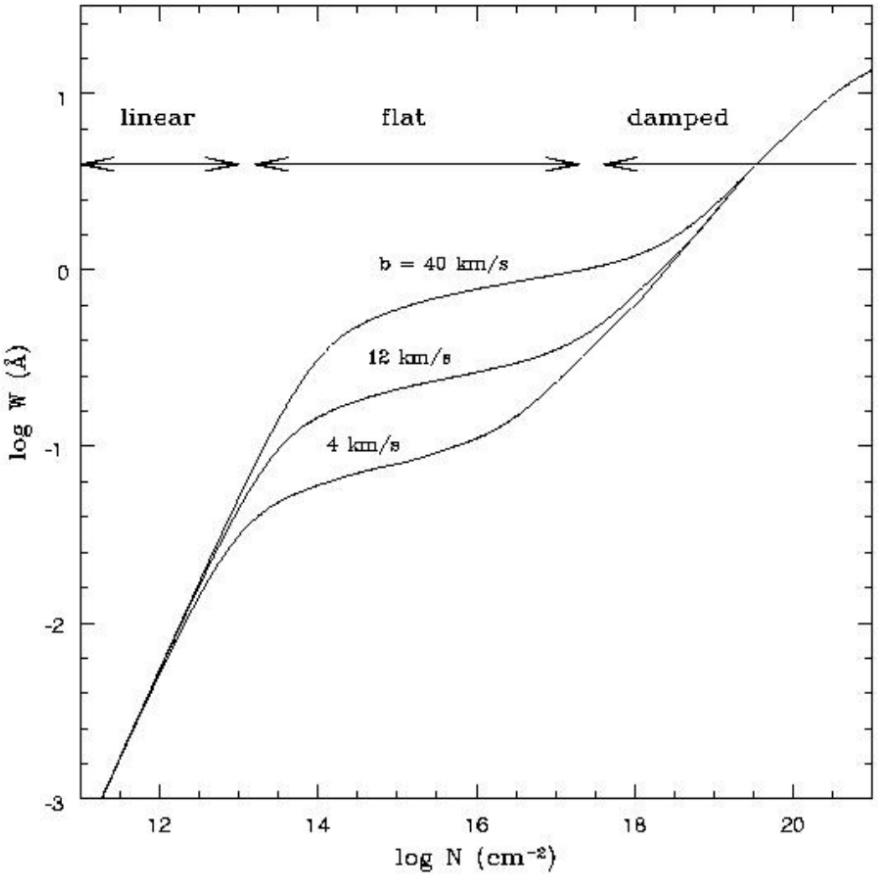

Fig. 2.2: Theoretical curve of growth for hydrogen Lyman α absorption.

For optically thin lines the EW is linearly proportional to $N_{\rm HI}$ or the oscillator strength f ($W(N) \propto Nf$) while for lines saturated in the Doppler cores $W(N) \propto ln(N)$ and for lines saturated in the Lorentzian wings $W(N) \propto \sqrt{N}$.

Most of the absorbers with $\log(N_{\rm HI}) < 14$ are optically thin in the Lyman-α transition and are thus on the linear part of the curve of growth. As such, their properties can be measured with a few (tenth) high-resolution ($\sim$ 10 km s^{-1}) high-signal to noise ratio ($S/N \sim 50$) spectra (Kim et al., 2002a).

At the highest column densities, $log(N_{\rm HI}) > 20.3$ (DLAs), the Lorentzian wings of the absorption line become obvious against the continuum of the background source. DLAs then can be identified and measured to high sta-

tistical precision with a large number of low to moderate resolution spectra of background QSOs, as it has been done using the QSO sample within the Sloan Digital Sky Survey (Noterdaeme et al., 2012). These absorbers are always associated with metal absorption lines, as shown in Figure 2.3.

Absorbers with intermediate column density values, $14 < log(N_{\rm HI}/atoms\,{\rm cm}^{-2}) < 20$, lie on the flat part of the curve of growth requiring specific techniques to be identified. LLSs (with $log(N_{\rm HI}) > 17$) have optical depth to hydrogen-ionizing photons $\tau \geq 1$, and may be recognized by the strong breaks they produce in background QSO spectra at the Lyman limit wavelength (912 Å restframe). As a consequence they can be easily discovered using low resolution spectra; however, large samples are needed because they are comparatively rare. For studies at $z \leq 2.6$, surveys of LLSs require space-based observations as the Lyman limit shifts below the atmospheric cutoff at UV wavelengths. However, over the last 30 years with a combination of ground and space based observations,the distribution of LLSs in the redshift range $2 < z < 3$ has been well characterized (Tytler, 1982; Sargent et al., 1989; Stengler-Larrea et al., 1995; Ribaudo et al., 2011; O'Meara et al., 2013; Prochaska et al., 2015). Finally there is the historically most-problematic column density range,$14 < log(N_{\rm HI}) < 17$. These intermediate-$N_{\rm HI}$ absorbers are saturated but have relatively low opacity and therefore weak "Lyman breaks". For this reason, putting observational constraints on their statistical incidence has been very challenging.

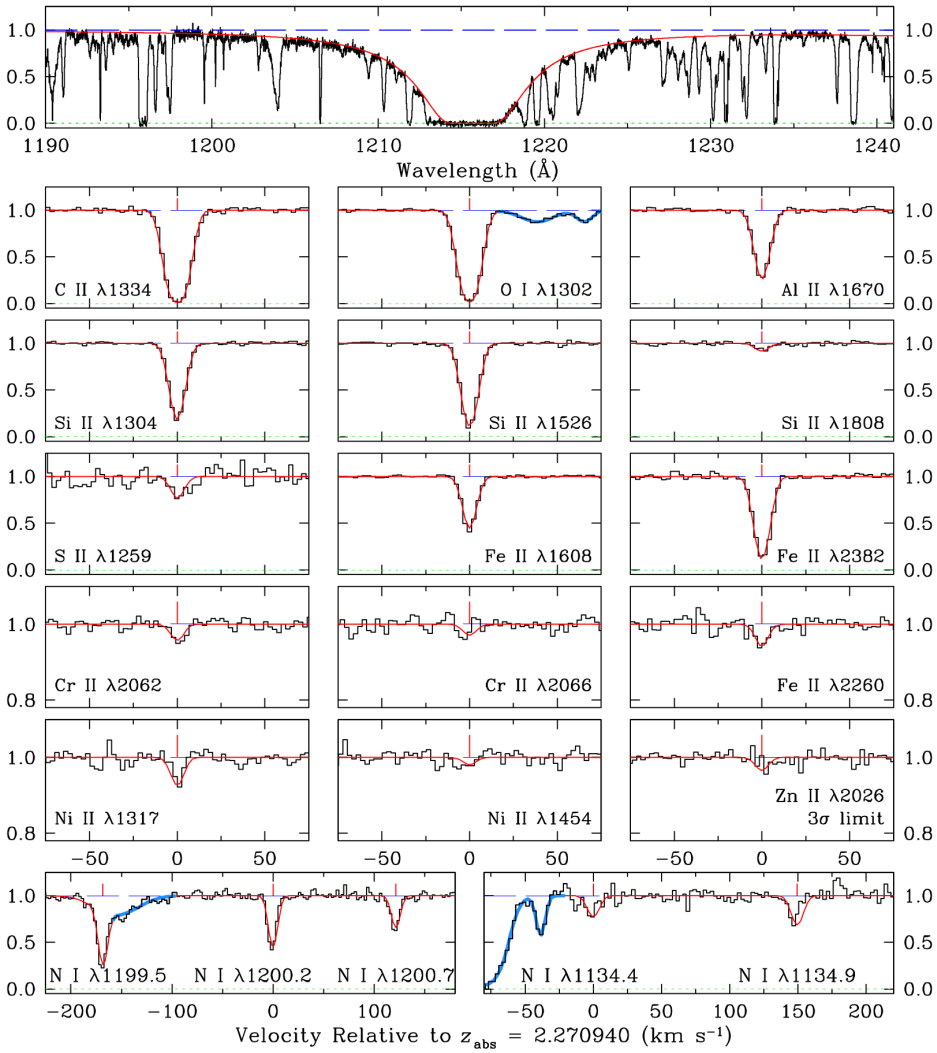

Fig. 2.3: The top panel shows the damped Ly α line for a metal-poor DLA at z_{abs} = 2.270940 towards the QSO J1111+1332. The continuous red line is the best-fitting Voigt profile model for an HI column density $log(N(HI)/cm^{-2}) = 20.39 \pm 0.04$. The remaining panels display a selection of the associated metal absorption lines, with the best-fitting model overplotted. Fitted line blends are shown by the blue lines. The red tick marks above the spectra indicate the position of the absorption component.Credits:fig. 8 by Cooke et al. (2015)

2.2 Lyman forest properties and evolution with redshift

The basic observational properties of the Lyman-α forest were established in the late 1970s and early 1980s, when the combination of 4-m telescopes (e.g. the AAT, KPNO, MMT, Palomar) and sensitive photon counting electronic detectors (e.g. the University College London's IPCS) first permitted quantitative spectroscopy of high redshift QSOs to be performed (Boksenberg et al., 1978; Young et al., 1979; Sargent et al., 1980). Making use of the new technology, the work by Sargent et al. (1980) set the stage for what for many years has been the standard picture of the Lyman forest: the absorption features in the forest were found to be consistent with a new class of astronomical objects, intergalactic gas clouds, which are distinct from galaxies (and metal absorption systems) by their large rate of incidence (or line density : dN/dz) and their weak clustering. Estimates of the gas temperature and the ambient UV flux are found to be consistent with a highly ionized optically thin gas kept at a temperature $T \sim 3 \times 10^4$ K by photo-ionization heating.

Initially it was assumed that the absorption features visible in quasar spectra were associated to pressure-confined clouds embedded in an even hotter inter-cloud medium (Sargent et al., 1980).

This model was abandoned with time since it was not capable of explaining some observations like the observed large range in column densities, and the evolution of the number of absorbers with redshift. Moreover, observations of the spectral shape of the cosmic microwave background, which are associated to a black-body, exclude the existence of an hot inter-cloud medium.

Thanks to new high-resolution and high-S/N data, especially from the Keck and VLT high-dispersion spectrographs, and to the breakthrough provided by numerical simulations in the late years of the XX century the picture globally accepted from the astrophysical community changed from "intervening galaxies / pressure confined IGM clouds" to the so-called *cosmic web*: the gas is arranged in filaments and sheets, closely tracing the dark matter distribution on large scales. In this model the low column density absorption systems are associated with the first large scale structure formed: sheet-like structures or pancakes of gas. The gas accretes through weak shocks and

settles in a dense, central cooling layer, presumably to form stars in some of the denser regions. At the lowest column densities gas remains unshocked and just bounces back because of the hydrostatic pressure. The gas is partly confined by dark matter gravity and partly by ram-pressure. Higher column density clouds arise in more filamentary structures, with column density contours of $N_{\rm HI} \sim 10^{14} cm^{-2}$ extending continuously and at relatively constant thickness ($\sim 40 - 100$ kpc proper) over Mpc distances. With increasing column density the absorber geometry becomes rounder. Such absorbers more closely correspond to the mini-halos/clouds mentioned previously; there the enclosed gas column density is high enough to make the absorption system appear as a Lyman limit or damped Lyman-α system.

Since the Universe expands approximately adiabatically and the Lyα forest is in photo-ionization equilibrium with the UV background, the temperature of the Lyα forest as a function of redshift provides a unique and powerful tool to probe the physical state of the IGM and the reionization history of the universe (Hui & Gnedin, 1997; Schaye et al., 1999; Ricotti et al., 2000; McDonald & Miralda-Escudé, 2001).

For a low-density ionized gas (at overdensity values $\delta < \sim 10$), the temperature of the gas is shown to be tightly correlated with the overdensity of the gas (see for details: Hui & Gnedin, 1997; Furlanetto & Oh, 2009), following the equation:

$$T = T_0 (1 + \delta)^{\gamma - 1} \tag{2.8}$$

where T is the gas temperature, T_0 is the gas temperature value at the mean gas density and $(\gamma - 1)$ is a constant at a given redshift z. Both T_0 and $(\gamma - 1)$ are a function of z, depending on the thermal history of the IGM (Hui & Gnedin, 1997).

This equation of state, however, is not directly observable. Instead of T and $(1 + \delta)$, observations only provide the neutral hydrogen column density $N_{\rm HI}$ (in atoms $\rm cm^{-2}$) and the Doppler parameter b (in km s^{-1}) of the forest absorption lines. In practice, a lower cutoff envelope in the $N_{\rm HI}$-b distribution is used to probe the upper limit on the temperature of the IGM, as done in Kim et al. (2002b)

Deriving $b_{\rm c}(N_{\rm HI})$ from observations depends on many factors, such as the method used for the line deblending, the number of available absorption lines, the metal-line contamination, and the method adopted for fitting the

lower $N_{\rm HI}$-b envelope (Hu et al., 1995; Kirkman & Tytler, 1997; Bryan & Machacek, 2000; McDonald et al., 2001; Ricotti et al., 2000; Schaye et al., 2000; Kim et al., 2002b). The different approaches and the limited numbers of lines have led, in part, to contradicting results on the evolution of $b_{\rm c}(N_{\rm HI})$ in the literature.

The simplest IGM property investigated with Lyman forest observations is the evolution with redshift of the line number density. In 2008 Saitta et al. (2008) derived the following evolution law:

$$\frac{dn}{dz} = (166 \pm 4)\frac{(1+z)}{3.5}^{(2.8\pm0.2)} \tag{2.9}$$

The evolution of the line density is governed by two physical mechanisms:

1. Hubble expansion
2. UV background

At high z the line density growth is steeper because the UV decreases with increasing redshift and the Hubble expansion blows recombination. At $z < 1.5$ the UV background starts decreasing with decreasing z and this implies a weaker evolution of the line density.

Strictly linked to the previous property is the evolution of the column density distribution function (CDDF) with reshift: Kim et al. (2002a) observed that the low column density of the CDDF becomes steeper at higher redshift while the high-column density part seems not to evolve very much.

Rudie et al. (2013) studying the evolution of the CDDF at low-intermediate column densities ($log(N_{\rm HI}) < 17$) derived the dependence of opacity on column-density.

The evolution of the effective opacity follows roughly a power-law from z $\sim$ 2 up to z $\sim$ 5:

$$\tau_{eff} = \tau_0 \left(\frac{1+z}{1+z_0} \right)^{\beta} \tag{2.10}$$

where $[\tau_0, \beta, C] = [0.751, 2.90, -0.132]$, for $z_0 = 3.5$.

With absorption lines observed in QSO spectra astronomers have constrained also the UV background intensity. To do it some assumptions are necessary: the assumption of spatial uniformity of the ionizing sources (which

implies an homogeneous UV background) and uncorrelated absorbing clouds of fixed column. The spatial uniformity of the background is justified at high redshift since the mean free path of ionizing photons, $\lambda_{\rm mfp}$, soon after the cosmological reionization becomes much longer than the average distance between sources, suppressing the variations at small scales (e.g Meiksin & White, 2004). Once made the two assumptions described above, the background specific intensity, I_ν , can be estimated with two inputs, the column density distribution of absorbing clouds, $\partial^2 N/\partial x \partial N_{\rm HI}$, where x is the comoving distance, and the physical specific emissivity of the sources, $\epsilon(x, \nu, z)$.

The intergalactic ionizing radiation field, a key ingredient in the development of reionization, is determined by the amount of ionizing radiation escaping from the host galaxies of stars and quasars. The value of the escape fraction as a function of redshift and galaxy mass remains a major uncertainty in all current studies, and could affect the cumulative radiation intensity by orders of magnitude at any given redshift. Romano et al. (2019) measured in their recent work an escape fraction equal to 0.49 at $z \sim 4$, in agreement with the values obtained for both brighter and fainter sources at the same redshift. They show that the mean free paths of ionizing photons are characterized by a skewed distribution function, with an average of $49 - 59$ proper Mpc. This value is larger than the one obtained at the same redshift by many authors in the literature using different techniques.

2.3 IGM metal enrichment

The observed intergalactic medium (IGM) shows substantial metal enrichment at all redshifts and at all column densities greater than $N_{\rm HI} = 10^{15} atoms\,{\rm cm}^{-2}$, from the Lyman-α absorbers at $z > 3$ with an average metallicity of 10^{-3} (Davé et al., 1998; Aracil et al., 2004; D'Odorico et al., 2016) to the DLA metallicity, which reaches values near to solar metallicities. The work by D'Odorico et al. (2016) show that at $z \sim 3$ the volume filling factor of the IGM gas enriched to a metallicity $log(Z/Z_\odot) \geq -3$ is of the order of $\sim 10\%$. These results favour a scenario in which metals are found also outside the CGM of bright star-forming galaxies, possibly due to pollution by lower mass objects and/or to an early enrichment by the first sources. A significant fraction of cosmic metal appears to lie in the IGM, and half of them lie in DLAs.

2.4 The importance of IGM for cosmology and galaxy formation studies

The study of the IGM is important for many astrophysical disciplines.

In cosmology, the IGM can be used to test our models of structure formation on the smallest comoving scales and viable alternatives to CDM (e.g. Viel et al., 2004; Viel, 2005; Seljak et al., 2006) and to understand the anisotropies in the cosmic microwave background (CMB; Ostriker & Vishniac, 1986; Hu, 2000). In addition, the same astrophysical processes that mould the IGM can bias cosmological parameter inferences from galaxy clustering (e.g. Pritchard & Furlanetto, 2007; Wyithe & Dijkstra, 2011).

The Lyman Forest offers a powerful tool to follow closely the distribution of Dark Matter (DM) down to the Jeans scale of the intergalactic gas, where the baryonic pressure counteracts gravity and causes a decoupling between DM and baryons, preventing the gas from following the spatial distribution of the DM (Viel, 2005).

In the galaxy formation field of study, the densest absorbers in the IGM (e.g LLSs and DLAs) provide information on galaxy evolution since they represent the reservoir from which galaxies take the gas that will be the fuel for the star formation (Rees, 1986; Efstathiou, 1992; Thoul & Weinberg, 1996; Kereš et al., 2005) and they are often associated with metal lines, which can be used to probe the enrichment of IGM and more in general the process of metal ejection in galaxies (Madau et al., 2001). The metal enrichment of IGM has become a very hot topic among the scientist who study the IGM. At very high redshift ($z >> 5$) population III stars could enrich the IGM to a low level (e.g. Carr et al., 1984; Ostriker & Gnedin, 1996; Haiman & Loeb, 1997; Tegmark et al., 1997; Abel et al., 1998), but since most metals presumably form in stars hosted in galaxies, the fundamental question at the basis of the study of the IGM metal enrichment is: how the first galaxies pollute effectively the IGM. To understand the enrichment of the IGM by galactic stars astrophysicists have focused on three mechanisms. First, metal-enriched gas (or stars that later explode as Type Ia supernovae) might be unbound during a merger or tidal interaction with another galaxy, or by the ram pressure of the IGM through which the galaxy moves. These processes in general are referred to as 'dynamical removal'. Second, the energy input from SNe may impart sufficient kinetic and thermal energy

to the gas in a galaxy to cause the gas escape from the gravitational well of the galaxy. This process is denoted as 'galactic wind mechanism'. Third, the radiation pressure on dust grains due to stellar light may exceed the gravitational force of the matter, leading to an outflow of dust (if the dust can decouple from the gas). This mechanism takes the name of 'radiation-pressure ejection' or 'dust ejection'.

In addition, the IGM can be used to measure the Ultra-Violet photon background (UVB) and to identify the sources of Reionization. The estimates of the UVB at different redshifts together with theoretical models for reionization provide a viable measurement of the total ionizing photon production of quasars and galaxies (Haardt & Madau, 1996; Miralda-Escudé, 2003; Faucher-Giguère et al., 2009; Becker & Bolton, 2013) as well as galactic stellar yields (e.g. Peeples et al., 2014). Thank to Lyman forest spectra we can estimate the mean free path of UV photons and the opacity of the IGM at different redshifts, which put constraints directly on the value of the UVB intensity and we can also study the IGM temperature evolution comparing it with theoretical model predictions to put independent constraints on the UVB topology and evolution (Becker et al., 2011; Boera et al., 2014; Bolton et al., 2014; Upton Sanderbeck et al., 2016).

The special interest in the observations of Lyman absorption features lies in their high sensitivity: with the present instrumentation (e.g Keck and VLT spectrographs) it is possible to detect neutral HI down to column density values of $10^{12} atoms^{-2}$ and up to $z \sim 7$, while for example 21 cm radio observations are limited in the best cases to column densities that are 6 orders of magnitude larger and confined to a much more local Universe.

In this way observations of the Lyman forest reveal very different structures at very different epochs of the Universe history, ranging from fluctuations or even under-densities of the diffuse intergalactic medium to the interstellar medium in protogalactic disks. One advantage of the Lyman-α observations is the relatively simple physics involved: often one can deal with optically thin gas and density fluctuations not far from the linear regime.

2.5 Observational approaches to the study of the IGM/galaxy connection

To investigate the connection between the densest absorbers in the IGM (DLAs) and galaxies the first essential instrument is the search for galactic emission in the vicinity of the HI absorption detected in quasar spectra. In the last twenty years, astronomers dedicated numerous follow-up observations to identify the counterparts of DLAs, mainly at low redshift (e.g. Chen & Lanzetta, 2003; Rao et al., 2011a; Rahmani et al., 2016). Different techniques have been used: narrow-band imaging of the fields around the background quasar (Møller & Warren, 1998; Kulkarni et al., 2007; Fumagalli et al., 2010; Rahmani et al., 2016), long-slit spectroscopy to search for emission lines from the galaxy associated with the DLA system (e.g. Møller et al., 2002; Fynbo et al., 2010, 2011; Noterdaeme et al., 2012; Srianand et al., 2016; Krogager et al., 2017), integral field spectroscopy (Péroux et al., 2011; Wang et al., 2015), sub-millimeter observations with ALMA (Neeleman et al., 2019), and wide field IFU spectroscopy provided by the Multi Unit Spectroscopic Explorer (MUSE) at the VLT (Bacon et al. 2010), working with adaptive optics (Mackenzie et al., 2019; Dutta et al., 2020). In the last year the first radio surveys, based on 21 cm asborption line detection, started: FLASH and MALS. These surveys will allow astronomers to constrain the properties of DLAs in the low-redshift Universe ($z < 2$). The connection between the densest absorbers in the IGM and galaxies is still poorly constrained by observations because the survey for galactic DLA counterparts have a low success rate in the identification of the counterparts (Fumagalli et al., 2015). This is interpreted as the result of the fact that DLAs are associated mostly to low SB galaxies, below the sensitivity of current instruments (Krogager et al., 2017) . The detection rate increases when strong cuts on the DLA metallicity are applied (Krogager et al., 2017). While in early DLA studies it was commonly believed that they originate from the absorption of gas settled in the disks of massive galaxies (Prochaska & Wolfe, 1997), there is now ample observational evidence that small and intermediate mass galaxies provide a not negligible contribution to DLA statistics (Krogager et al., 2017), in accordance with the predictions of the theoretical studies by Rahmati & Schaye (2014) and Pehlivan Rhodin et al. (2019). The recent discovery of strong HI absorbers detected at high impact parameters

($b > 30$kpc) from their likely host galaxies (Christensen et al., 2019; Møller & Christensen, 2020; Péroux et al., 2019; Hamanowicz et al., 2020) provides insight into their origin and clustering properties. As observed by Hamanowicz et al. (2020) the galaxies with the smallest impact parameters are not necessarily the closest to the absorbers in velocity space.

2.6 Theoretical approaches to the study of the IGM/galaxy connection

The first theoretical studies on the IGM were based on analytical models and cosmological simulations (Gardner et al., 1997, 2001; Haehnelt et al., 1998; Prochaska & Wolfe, 1997; Wolfe & Prochaska, 1998). Since the 1997, different groups developed hydro-dynamical simulations devoted to the study of cold gas in/outside galaxies which shed light on the IGM origin and evolution. The principal ones are: EAGLE (Schaye et al., 2015; Crain et al., 2015, 2017), ILLUSTRIS-TNG (Vogelsberger et al., 2014; Weinberger et al., 2017; Pillepich et al., 2018), FIRE (Hopkins et al., 2014, 2018). Up to now it has been very difficult to study the physical properties of all the classes of absorbers associated to the IGM with an unique simulation because of the large dynamical range required for this analysis, that leads to a very high computational cost. The studies of the IGM in its entirety have been carried out using cosmological hydro-dynamical simulations, with intermediate/low resolution (Gardner et al., 1997, 2001; Haehnelt et al., 1998; Cen et al., 2003; Nagamine et al., 2004) while high-resolution hydro-dynamical simulations with small boxes or zoom-in simulations have been adopted for the study of the denser absorbers (LLS and DLAs) (Nagamine et al., 2004; Pontzen et al., 2008; Tescari et al., 2009; Razoumov, 2009; Fumagalli et al., 2011; Cen, 2012; van de Voort et al., 2012; Altay et al., 2013; Pehlivan Rhodin et al., 2019; Hassan et al., 2019).

The semi-analytic method has been used historically to address the IGM enrichment in greatly varying levels of complexity and in more recent works to study the physical origin of DLA systems (Berry et al., 2014; Theuns, 2021). In particular, Nath & Trentham (1997); Gnedin & Ostriker (1997); Ferrara et al. (2000), Madau et al. (2001) and Fumagalli et al. (2015) have studied IGM enrichment by winds in this way, attempting also to calculate the statistical properties of the metal distribution in the IGM. Cen & Ostriker (1999)

and Gnedin & Ostriker (1997) have used SAMs to study IGM enrichment by dynamics or other processes.

In the next two sections I will introduce how hydro-dynamical simulations and semi-analytic models are built and what are the advantages/disadvantages of each tool for the study of DLA systems and the cold gas cycle in galaxies.

2.6.1 Hydro-dynamical simulations

Historically the schemes adopted in hydro-dynamical simulations to treat the gas dynamics were of two types: Lagrangian scheme and Eulerian schemes. In the Lagrangian schemes the fluid elements are treated as particles and the equations describe the evolution, in space and time, of the properties of the fluid particles. This scheme is similar to the one adopted for the DM dynamics and it's the easiest to implement. Unfortunately it has been shown that this scheme cannot properly model fluid mixing and sub-sonic turbulence (Agertz et al., 2007; Springel, 2010a) and have problems in treating shocks, which can be captured only by adding to the equations an artificial viscosity. The most common Lagrangian scheme is the Smoothed Particle Hydrodynamics (SPH), adopted since the first 3D simulations (Efstathiou & Eastwood, 1981; Evrard, 1988; Hernquist & Katz, 1989). It evaluates the value of a generic variable at any point as a smoothed estimate, summing over nearby particles using the kernel function within the smoothing length h. Due to its adaptive spatial resolution and good conservation properties this scheme has been very popular for galaxy formation simulations until today (Gingold & Monaghan, 1977; Springel, 2010a; Somerville & Davé, 2015). Current SPH codes are GADGET (Springel et al., 2005) , GASOLINE (Wadsley et al., 2004), VINE (Wetzstein et al., 2009), HYDRA (Couchman et al., 2011), CHANGA (Menon et al., 2015), SWIFT (Theuns et al., 2015) and GIZMO (Hopkins, 2014) . Most of the recent SPH codes include updated implementations to treat the mixing problem better (see e.g. Hopkins et al., 2014; Hu et al., 2014; Read & Hayfield, 2012; Schaller et al., 2015; Schaye et al., 2015; Wadsley et al., 2008, and references therein). A valid alternative to the SPH method particle based are the meshless-finite-mass and meshless-finite-volume proposed by Gaburov & Nitadori (2011). The most recent SPH implementation, GIZMO, is based on the GADGET framework and shows some significant improvements on idealised test problems, in particular for

low Mach number gas (Hopkins, 2015). Eulerian hydrodynamic codes have also been widely used for cosmological simulations, some with adaptive mesh refinement capabilities. These codes typically perform better than SPH in terms of mixing and shock problems but might suffer from artifacts due to grid structure and numerical diffusion. The first Eulerian treatment was designed by Cen & Ostriker (1992) and the most used Eulerian adaptive mesh refinement codes are ENZO (Bryan et al., 2014), RAMSES (Teyssier, 2002), ART (Kravtsov et al., 1997) and also FLASH (Fryxell et al., 2000) as well as ATHENA (Stone et al., 2008) for ISM simulations on smaller scales.

The weaknesses of SPH and AMR codes lead to the need of developing novel numerical schemes more accurate in the regime relevant for cosmic structure formation. Springel (2010b) proposed a novel scheme based on a moving unstructured mesh defined by the Voronoi tessellation of a set of discrete points. In this scheme the mesh-generating points can in principle be moved arbitrarily. If they are chosen to be stationary, the scheme is equivalent to an ordinary Eulerian method with second order accuracy. If they instead move with the velocity of the local flow, one obtains a Lagrangian formulation of continuum hydrodynamics that does not suffer from the mesh distortion limitations. The newly developed code AREPO (Springel, 2010b) combines advantages of the Lagrangian and Eulerian approaches and performs much better than traditional SPH codes, like GADGET, on mixing problems with a high convergence rate (Sijacki et al., 2012; Springel, 2010b). There are ongoing efforts to better understand the strengths and weaknesses of different numerical schemes (e.g. Hayward et al., 2014; Heitsch et al., 2011; Hubber et al., 2013; Kim et al., 2014; Price & Federrath, 2010; Naab & Ostriker, 2017) and to constantly improve on accuracy and performance of all major codes.

It has been realized early on that different numerical schemes applied to cosmological simulations can result in systems with different physical properties (Frenk et al., 1999), even if only gravity and hydrodynamics are considered. In addition, there is a wealth of published sub-resolution models which are used to model galaxy formation. These models are often designed for particular numerical schemes and introduce even stronger variations in physical properties for a given set of initial conditions (Scannapieco et al., 2012). One of the major challenges in computational galaxy formation is to further improve on the numerical schemes and reduce the

contribution of sub-resolution modeling to numerically resolved physical scenarios. Modern cosmological hydrodynamical simulations typically include a sub-resolution model to describe the star formation process (e.g. Ascasibar et al., 2002; Springel & Hernquist, 2003; Dubois & Teyssier, 2008a; Schaye & Dalla Vecchia, 2008; Few et al., 2012), radiative cooling (e.g. Katz et al., 1996; Wiersma et al., 2009a) and the chemical enrichment of the ISM and CGM (e.g. Steinmetz & Mueller, 1994; Mosconi et al., 2001; Lia et al., 2002; Springel & Hernquist, 2003; Kobayashi, 2004; Scannapieco et al., 2005; Tornatore et al., 2007; Oppenheimer & Davé, 2008; Wiersma et al., 2009b; Few et al., 2012). Furthermore, some small-scale simulations also include a specific treatment for magnetic fields (e.g. Teyssier, 2002; Dolag & Stasyszyn, 2009; Pakmor et al., 2011; Pakmor & Springel, 2013; van de Voort et al., 2019), radiative transfer (e.g. Abel & Wandelt, 2002; Cantalupo & Porciani, 2011; Petkova & Springel, 2011a,b; van de Voort et al., 2019), cosmic ray physics (e.g. Jubelgas et al., 2008) and thermal conduction (Dolag et al., 2004; Jubelgas et al., 2004; Arth et al., 2014). Most importantly, hydrodynamic simulations of galaxy formation usually include prescriptions for the stellar feedback (e.g. Dekel & Silk, 1986; Navarro & White, 1993; Mihos & Hernquist, 1994; Gerritsen & Icke, 1997; Thacker & Couchman, 2000; Kawata & Gibson, 2003; Sommer-Larsen et al., 2003; Springel & Hernquist, 2003; Brook et al., 2004; Oppenheimer & Davé, 2006; Scannapieco et al., 2005; Stinson et al., 2006, 2013; Dalla Vecchia & Schaye, 2008; Dubois & Teyssier, 2008b; Okamoto et al., 2010; Piontek & Steinmetz, 2011; Simpson et al., 2015), and in some cases also for the AGN feedback (e.g. Di Matteo et al., 2005; Springel et al., 2005; Kawata & Gibson, 2005; Sijacki & Springel, 2006; Thacker et al., 2006; Sijacki et al., 2007; Okamoto et al., 2008a; Booth & Schaye, 2009; Kurosawa et al., 2009; Debuhr et al., 2011; Teyssier et al., 2011; Dubois et al., 2012; Choi et al., 2020).

In 1998, Haehnelt et al. (1998) showed that while the velocity width profiles of DLAs can be reproduced by rotation in discs as well as proto-galactic clumps,the velocity width distribution cannot be reproduced considering the rotation velocity of gas in discs in the context of a cold dark matter (CDM) model.

Nagamine et al. (2004) focused on the effects of star formation and super-nova (SN) feedback on DLAs that were largely neglected by the earlier numerical works (e.g. Gardner et al., 1997, 2001) and showed that the

distribution of DLAs could be significantly affected by SNe feedback. Razoumov (2009) studied the effects of radiative transfer by post-processing the results of cosmological hydrodynamic simulations, but their simulations did not include models for star formation and SN feedback. It was difficult at that time to treat the radiative transfer dynamically from all galaxies in a large volume of space, as well as star formation and feedback by SNe and active galactic nuclei in a cosmological box, due to the heavy computational cost required for such a simulation with full treatment of all physical processes. Therefore, simulators usually limited their studies to a small number of objects, in small/intermediate boxes, in order to treat radiative transfer self-consistently with the star formation in simulations. In 2007 Nagamine employed a series of simulations with different resolution, box sizes (from $L_{box} = 4$ cMpc to 100 cMpc) and feedback strengths to study the effects of galactic wind feedback on the distribution of DLAs. They focused on two scenarios: 'no wind' feedback and 'strong wind' feedback, corresponding to the wind speed of $v_w = 0$ and 484 km s^{-1}, respectively. They found that increasing the strength of galactic wind feedback,the fraction of DLAs in low-mass halos is reduced, and a larger fraction of DLAs will be in more massive halos that are farther away, resulting in a broader b-parameter distribution. In 2008 Pontzen et al. (2008) developed a SPH simulation with a model incorporating supernova feedback, via galactic winds, and simple radiative transfer, and studied the comparison between the simulation predictions and the observed DLA metallicity and abundance. They found that the haloes with virial mass $10^9 < M_{200} < 10^{11}\mathrm{h}^{-1}\mathrm{M}_\odot$ were the main contributors to the DLA cross-section. In the same year the work by Barnes & Haehnelt (2009), based on a semi-analytical model, lead to a different result. They claimed that in order to reproduce the velocity widths distribution of DLAs the contribution of haloes less massive than $10^{10}\mathrm{h}^{-1}\mathrm{M}_\odot$ should be suppressed.

In 2009 Tescari et al. (2009), with the help of a hydro-dynamical simulation based on GADGET2, investigated the impact of galactic feedback, in the form of outflows, on the DLAs properties. Their fiducial simulation run employed a strong wind (SW), and the following parameters: 10 $h^{-1}Mpc$ box size (comoving), 320^3 particles, mass resolution $3.5 \times 10^5 h^{-1} M_\odot$ and $1.5h^{-1}$kpc (comoving) gravitational softening length. They found that strong winds are requested to reproduce the observed number density of DLA at high redshift. According to Tescari et al. (2009), the disc-like model of

Prochaska & Wolfe (1997), in which DLAs are thick rotating discs with speed typical of a present-day spiral galaxies, and the alternative model based on the assumption that DLAs are proto-galactic clumps (Haehnelt et al., 1998) seem to be both viable.

In the first DLA studies based on hydro-simulations (e.g. Gardner et al., 1997; Haehnelt et al., 1998; Nagamine et al., 2004), self-shielding was modelled with 1D radiative transfer calculations or correlations between neutral column density and total volume density but in 2007 computing hardware and algorithms have advanced to the point where it was feasible to incorporate full 3D radiative transfer to calculate self-shielding (Kohler et al., 2007; Pontzen et al., 2008; Altay et al., 2011; Fumagalli et al., 2011; McQuinn et al., 2011; Cen, 2012). Faucher-Giguère & Kereš (2011); Altay et al. (2011); Yajima et al. (2012) and Rahmati et al. (2013) studied in detail the effects of radiative transfer on DLA self-shielding, finding that taking into account properly radiative transfer impact the spatial distribution of DLAs, and their ionization state. Another important process to be considered in theoretical DLA studies is the conversion of atomic to molecular hydrogen, in the ISM. Even if the conversion of HI to H_2 has been thought to determine the high end cutoff of the DLA CDDF (Schaye, 2001; Krumholz et al., 2009a) the molecular hydrogen formation was considered only by few simulations (e.g Cen, 2010; Altay et al., 2011; Erkal et al., 2012) when modeling the HI absorption.

Different DLA studies were based on the simulations in the OWLS project: Altay et al. (2013) focused on the study of the column density distribution within these models, while Rahmati et al. (2013) looked at the typical host halos of DLAs and LLSs, and one of the simulations most used in the field of DLA studies nowadays (e.g Garratt-Smithson et al., 2021) originated from the same project (EAGLE, Schaye et al., 2015)

In the same year also also the group guided by R. Davè developed a hydro-dynamical simulation suite devoted to DLA studies. Davé et al. (2013) and Bird et al. (2014, 2015) used a similar feedback scheme and examined the galactic HI mass function and DLA CDDF respectively.

Similar studies have been done with grid-based codes (Fumagalli et al., 2011; Cen, 2012). Cen (2012) was one of the few studies that investigated the relation between large-scale environment and DLA statistics. This work is based on two simulations, that probe one overdense and one underdense

region, respectively, addressing different observed properties of DLAs, from the kinematic distribution of metal lines to the column density distribution function.

So far we have seen the results obtained with SPH and grid-based codes but also the AMR scheme has been adopted in simulations developed for DLA studies: Razoumov (2009) examined the kinematic distribution of DLA metal lines in adaptive mesh refinement (AMR) simulations of isolated haloes.

In 2015 the work by Bird et al. (2015) has shown that feedback processes are important for the properties of the DLAs with highest N_{HI}.

In the last decade several hydro-dynamical simulations addressed the topic of the DLA/galaxies connection, usually focusing on some of the observed DLA properties and getting often conflicting results regarding the origin of these systems (Bird et al., 2015; Pehlivan Rhodin et al., 2019; Hassan et al., 2019).

In particular, Hassan et al. (2020) showed that the DLA abundance and metallicity distribution are sensitive to the galactic outflows' feedback scheme and the UVB amplitude adopted in the simulation, with a detailed comparison of the prediction by two state-of-the-art cosmological hydrodynamic simulations, SIMBA and TECHNICOLOR DAWN, while Pehlivan Rhodin et al. (2019), with a study based on cosmological hydrodynamic zoom simulations, found out that at $z > 1$ the halo plays an increasingly important role in the sub-DLA/DLA covering fraction, while at $0.4 < z < 1$ the disc and halo contribute with $\sim 60(80)$ and $\sim 40(20)$ per cent to column densities above the sub-DLA(DLA) lower limits.

The hydro-dynamical simulations have proven to be good tools for DLA studies because they can trace the kinematics and spatial distribution of cold gas in different galactic environments (e.g. ISM, CGM and inflows/outflows) but they cannot easily disentangle the physical origin of the cold gas properties at intermediate column density from resolution effects and numerical artefacts. Many works (e.g Diemer et al., 2018; Kauffmann et al., 2019) show that the nature and evolutionary effects of simulated baryon cycles depend heavily on the galaxy formation model, the numerical implementation, and the codes employed in numerical simulations. Therefore, we need to complement the study of the gas cycle in galaxies based on hydro-simulations, adopting semi-analytical models, in order to test different prescriptions for the sub-grid physics with a low computational cost.

2.6.2 Semi-analytical models

Semi-analytic Models (SAM) combine an analytic treatment of the physical processes driving galactic evolution, with halo histories contained in the halo merger trees. These models attempt a direct treatment of the physical processes driving the evolution of the baryonic component in galaxies. The treatment at the basis of each SAM consists of a collection of analytic prescriptions regulating the interactions of baryons in galaxies. These prescriptions are derived from observational relations or from dedicated high-resolution hydro-dynamical simulations, in a similar way to the sub-grid physics implemented in cosmological hydro-dynamical simulations. The halo merger trees provide the skeleton for the SAM analytic treatment,allowing galaxy evolution to be followed in a cosmological context. The merger trees can be constructed from a cosmological N-body simulation, or using Monte-Carlo methods based on the Press-Schechter formalism, or its later modifications. Adopting analytic merger trees offers the advantage of high resolution at small computational cost but does not take into account the observed assembly bias, which relates the DM halo growth to large-scale environment. The direct comparison with results of a full N-body simulation highlights the limits of this class of methods, and several corrections have been proposed to overcome these limits (see for example Parkinson et al., 2008; Neistein & Dekel, 2008) The first SAMs were created by White & Frenk (1991); Cole (1991); Lacey & Silk (1991). They included prescriptions for gas cooling, star formation and stellar feedback and were applied to halo distributions at various redshifts, calculated by means of analytic models for the collapse of the density fluctuation field. We have to wait the 1997 to see the first SAMs coupled to merger trees extracted from cosmological N-body simulations (Roukema et al., 1997; Kauffmann et al., 1999).
Nowadays the majority of SAMs is coupled to DM-only N-body simulations (Kravtsov et al., 2004; Croton et al., 2006; De Lucia et al., 2014).

In the last 20 years, SAMs have been updated and extended to investigate more in detail the various aspects of galaxy formation and evolution (see discussion in the reviews by Baugh, 2006; Benson, 2010). Thanks to their limited computational cost and large associated box volume, the semi-analytical models are very useful to study the evolution and the statistics of galaxy properties in cosmological volumes and they are the ideal tool for

making mock catalogs to compare with galaxy surveys. At the meantime we can obtain interesting predictions on DLA properties from SAMs, which is the topic of this PhD thesis. In my PhD activity I have analysed the properties predicted by the state-of-the-art semi-analytical model GAEA (De Lucia et al., 2014; Hirschmann et al., 2016; Xie et al., 2017) for simulated DLAs in the redshift range $2 < z < 3$. The first results of this study is described in details in Chapter 3, while in Chapter 4 I will discuss the modifications of the prescriptions regarding the Reionization history and feedback, the cold gas partitioning, and the treatment of the ejected gas, that I have introduced in the model in order to get a better agreement with the DLA observations at intermediate redshift.

2.6.3 Cold gas partitioning in SAMs and simulations

Since the derived properties for the simulated DLA, both in hydro-dynamical simulations and SAMs, depend strongly on the modelled distribution of HI in simulated galaxies, it is important to discuss the adopted cold gas partitioning both in simulations and in semi-analytical models. Until 10 years ago, most of the simulations and semi-analytical models do the partitioning of the cold gas in post-processing, not considering in this way the relation between the galaxy molecular content/distribution and the SF activity, and this is still the case for almost all the hydro-dynamical simulations, exceptions are SIMBA (Davé et al., 2019) and the simulation by van de Voort et al. (2019). Moreover, the partitioning procedure followed in simulations and semi-analytical models differs, on average, because in the post-processing analysis of the hydro-dynamical code results the ingredient of the partitioning recipes is the 3D gas density while in semi-analytical models it is the 2D density, and this leads to a discrepancy in the molecular/atomic ratio assigned to simulated galaxies in different stellar mass bins (e.g. Diemer et al., 2019; Stevens et al., 2019).

All the more up-to-date semi-analytical models for galaxy formation (e.g. GAEA, the SC model) assume a partitioning scheme observationally or theoretically motivated (e.g BR, KMT) and derive in a self-consistent way the HI/H_2 content of modeled galaxies, and in some cases also the spatial distribution of the molecular and atomic gas over the entire galactic disk (e.g GAEA-X17).

In the past this was not the case and two methods have been adopted to assign the atomic/molecular content to simulated galaxies:

1. Obreschkow & Rawlings (2009) firstly estimate the molecular ratio $R_{mol} = M_{\rm HI}/M_{\rm H2}$ and therefore assign the HI mass to model galaxies. To estimate R_{mol} they assume the following relation: $R_{mol} = [3.44{R_c}^{-0.506} + 4.82{R_c}^{-1.054}]^{-1}$, where $R_c \sim [{r_d}^{-4} M_{CG}(M_{CG} + 0.4M_\star)]^{-0.8}$ derived combining the gas density profile of galaxies observed in the Local Universe (Leroy et al., 2008a) with the relation between the molecular fraction and the mid-plane pressure (Blitz & Rosolowsky, 2006).

2. Baugh et al. (2004) adopt instead a simple conversion factor from cold gas mass to HI mass. They first divide the cold gas into hydrogen and

helium, adopting the universal fraction value of 24% for Helium+metals, and assuming that the ionized fraction of hydrogen in the ISM is negligible. The remaining cold gas in galaxies is therefore divided into $\sim 71\%$ of HI and 29% of H_2, leading to a universal value for the molecular ratio, equal to $R_{mol} \simeq 0.4$ (in agreement with observations by Keres et al. (2003); Zwaan et al. (2005)).

Power et al. (2010) have tested the impact on galaxy properties of these two methods for the derivation of R_{mol}, in 4 different semi-analytical models, finding little differences at low z and larger differences at high z.

The first SAM which has implemented a cold gas partitioning scheme, in a self-consistent way, was the one developed by Fu et al. (2010). After this model also Lagos et al. (2011) and Somerville et al. (2015) implement a self-consistent treatment of the cold gas partitioning in the ISM of model galaxies. Nowadays almost all the semi-analytical models implement detailed schemes for the cold gas partitioning (e.g. SHARK, L-GALAXIES, SANTA-CRUZ, GAEA, described in Lagos et al., 2018; Henriques et al., 2020; Xie et al., 2017; Yung et al., 2019; Popping et al., 2019a) It is important to apply the cold gas partitioning on the fly to derive self-consistently the effects on the star formation and gas accretion in galaxies(e.g. in many models the star formation rates depends directly on the surface density of the molecular hydrogen in the ISM, and not on the cold gas surface density).

In particular, for the GAEA model, as described in Xie et al. (2017), it has been implemented an explicit treatment for the partition of cold gas into its atomic (HI) and molecular components (H_2), and a star formation law based on the molecular abundance.

Chapter 3

DLA and HI in galaxies

This chapter is based, with limited modifications, on the published paper Di Gioia et al. (2020).

3.1 Introduction

With present and upcoming facilities (e.g. MUSE, ALMA, ELT), allowing us to trace the gaseous components of galaxies out to their outskirts, this is an ideal time to study the cycle of gas (and metals) in and around galaxies. Hydrogen is the most abundant element in the galactic cold phase, and can be detected in emission (21cm line - mostly in the local Universe) or in absorption (Lyman-α line, in the optical for $z \geq 1.65$ and in the UV for lower redshifts). Due to the sensitivity of current instrumentation, the detection in emission (21cm line) is strongly biased towards the brightest galaxies/highest column densities, and is limited to relatively low redshift (up to $z = 0.06$). In the last decade, the HI content of galaxies has been characterized for a large sample of local galaxies thanks to surveys like HIPASS, ALFALFA, GASS (Meyer et al., 2004; Giovanelli et al., 2005; Catinella et al., 2010, 2013, 2018). These surveys have also allowed studies of the correlation between HI and galaxy stellar mass or other galaxy properties (e.g. star formation rate, environment, etc.).

Studies based on absorption lines are not affected by the same observational limits of emission line studies: the Lyman-α line results from a transition between the 2^2 P state and the 1^2 S (ground) state of the hydro-

gen atom ($\lambda = 1215$ Å), and it is possible to observe it from the ground at z$\gtrsim 1.6$. The systems characterized by the strongest absorption lines are the Damped Lyman-α systems (DLAs), defined as hydrogen absorbers with column density $N_{\rm HI} > 10^{20.3}$ atoms cm^{-2}. These strong absorbers are typically associated with low-ionization metal line complexes (Prochaska et al., 2003; Noterdaeme et al., 2012; Rafelski et al., 2012), suggesting that they are part of a gaseous medium affected by chemical enrichment, like the ISM and the CGM in galaxies.

Large spectroscopic surveys, such as the Sloan Digital Sky Survey (SDSS; Schneider et al. 2010) and BOSS (Eisenstein et al., 2011), have greatly improved the statistics for samples of high-redshift absorbers ($1.5 < z < 4.5$), tightening the constraints on the shape of the column density distribution function, the comoving line density of DLAs, and the evolution of the neutral gas density (e.g Storrie-Lombardi & Wolfe, 2000; Péroux et al., 2003; Noterdaeme et al., 2012; Crighton et al., 2015). These studies have demonstrated that DLAs contain $\sim 80\%$ of the neutral gas available for star formation (Prochaska & Wolfe, 2009; Noterdaeme et al., 2012; Zafar et al., 2013; Storrie-Lombardi & Wolfe, 2000; Péroux et al., 2003; Prochaska et al., 2005), so DLAs studies provide us with an estimate of the gas available for star formation from $z = 5$ to now.

Rafelski et al. (2012) and Neeleman et al. (2013) have estimated the metallicities for a sample of DLAs in the redshift interval $(2 < z < 4)$, and investigated their mean metallicity evolution. Recently, these measurements have been updated by De Cia et al. (2018) who developed a procedure to estimate DLA metallicities corrected for dust depletion.

In the last decades, numerous follow-up observations have been carried out to identify the counter-parts of DLAs, mainly at low redshift (e.g. Chen & Lanzetta, 2003; Rao et al., 2011a; Rahmani et al., 2016). Different techniques have been used: narrow-band imaging of the fields around the background quasar (Møller & Warren, 1998; Kulkarni et al., 2007; Fumagalli et al., 2010; Rahmani et al., 2016), long-slit spectroscopy to search for emission lines from the galaxy associated with the DLA system (e.g. Møller et al., 2002; Fynbo et al., 2010, 2011; Noterdaeme et al., 2012; Srianand et al., 2016; Krogager et al., 2017), integral field spectroscopy (Péroux et al., 2011; Wang et al., 2015), and sub-millimeter observations with ALMA (Neeleman et al., 2019). The detection rate in blindly selected samples remains very

low (Fumagalli et al., 2015), but increases when strong cuts on the DLA metallicity are applied (Krogager et al., 2017). These results suggest that DLAs are likely associated with low-luminosity galaxies, most of which are below current observational capabilities (Krogager et al., 2017).

The occurrence of strong HI absorbers detected at high impact parameters ($b > 30$kpc) from their likely host galaxies (Christensen et al., 2019; Péroux et al., 2019) provides insight into their origin and clustering properties. While in early DLA studies it was commonly believed that they originate from the absorption of gas settled in the disks of massive galaxies (Prochaska & Wolfe, 1997), there is now ample observational evidence that small and intermediate mass galaxies provide a non negligible contribution to DLA statistics (Krogager et al., 2017), in accordance with the predictions of the theoretical study by Rahmati & Schaye (2014).

Font-Ribera et al. (2012) carried out a cross-correlation analysis of DLAs (selected from the BOSS survey) with the Lyman-α forest and obtained constraints on the DLA cross-section as a function of halo mass. The bias they find implies a typical DLA host halo mass of $\sim 10^{12} M_{\odot}$ at $z = 2$. In 2018 Pérez-Ràfols et al. (2018a) updated the results by Font-Ribera et al. (2012) finding a typical DLA halo mass of $\sim 4 \times 10^{11} M_{\odot}$. In the meantime, Arinyo-i-Prats et al. (2018) developed a new method to classify the metal strength of DLAs and studying the dependence of the bias on the metallicity of the absorbers Pérez-Ràfols et al. (2018b) showed that the linear bias associated with DLAs decreases as their metallicity decreases.

In the last 20 years, a number of theoretical studies have used hydrodynamical simulations to investigate the nature of strong HI absorbers and DLAs in particular (e.g. Gardner et al., 1997, 2001; Haehnelt et al., 1998; Nagamine et al., 2004; Pontzen et al., 2008; Tescari et al., 2009; Razoumov, 2009; Fumagalli et al., 2011; Cen, 2012; van de Voort et al., 2012; Altay et al., 2013; Rhodin et al., 2019; Hassan et al., 2020).

The resolution of the simulations has increased over time, but the approach typically needs to resort to different layers of sub-grid prescriptions to model the high HI column densities of DLAs. Some studies overcome the absence of a full cosmological distribution of absorbers by combining results from small-scale simulations with analytic parametrizations of the halo mass function to predict statistical properties of the DLA population (e.g. Gardner et al., 1997, 2001), or to study the nature of the host galaxies (e.g.

Pontzen et al., 2008). This approach can lead to biased results, requiring some strong assumptions about the environments that can give rise to DLA absorbers. In addition, it does not account for the potentially large scatter in the distribution of absorbers for haloes of similar properties.

Studies based on hydro-dynamical simulations have pointed out an important contribution to the DLA population, typically increasing with increasing redshift, from gas that is not associated with the ISM of galaxies. There is no consensus on the quantitative estimate of such a contribution that ranges, depending on the study, between ~ 20 per cent (Villaescusa-Navarro et al., 2018) to more than ~ 50 per cent (Fumagalli et al., 2011; van de Voort et al., 2012). Most numerical studies indicate a major contribution to the DLA population at $2 \leq z \leq 3$ from haloes with virial masses of $10^{10} - 10^{12} M_{\odot}$ (Cooke et al., 2006; Pontzen et al., 2008; Barnes & Haehnelt, 2009; Font-Ribera et al., 2012).

In this study, we focus on an alternative theoretical approach provided by semi-analytic models of galaxy formation. While these are unable to resolve the internal structure of galaxies and do not model the hydro-dynamical processes self-consistently, they can easily access to much larger cosmological volumes than hydrodynamical simulations. In addition, a fast exploration of the parameter space and an efficient investigation of the influence of different specific assumptions are possible, thanks to the limited computational costs. We take advantage of the state-of-the-art semi-analytic model GAlaxy Evolution and Assembly (GAEA, De Lucia & Blaizot (2007); De Lucia et al. (2014); Hirschmann et al. (2016)), coupled to large cosmological N-body simulations, and analyse the properties of host DLA galaxies, as well as their connection with dark matter haloes. GAEA accounts for an explicit partition of the cold gas between atomic and molecular hydrogen, but assumes that all cold gas is associated with galaxy disks. Our approach therefore ignores the contribution to DLAs from filamentary structures or gas outflows, and tests to what extent current estimates of DLA statistics can be explained by the gas in galaxy disks.

The specific questions that we want to address in our study include:

- What is the typical virial mass of dark matter haloes hosting DLAs?
- To what extent can we reproduce the observed DLA statistical properties, by only considering the ISM associated with galaxies?

- What drives the evolution of $\Omega_{\rm DLA}$ with z, and what is the contribution to this quantity of galaxies with different mass?

3.2 Simulated galaxies

3.2.1 The N-body Simulations

The adopted physical model for the evolution of galaxies and their baryonic components is coupled to the output of cosmological dark matter simulations, as detailed in De Lucia & Blaizot (2007). In this study, we use dark matter merger trees from two cosmological N-body simulations: the Millennium simulation (MSI; Springel et al. 2005), and the Millennium II simulation (MSII; Boylan-Kolchin et al. 2009).

Both the MSI and the MSII assume a WMAP1 cosmology, with $\Omega_m = 0.25$, $\Omega_b = 0.045$, $\Omega_\lambda = 0.75$, $h = 0.73$ and $\sigma_8 = 0.9$. Recent measurements from Planck (Planck Collaboration et al., 2016) and WMAP9 (Bennett et al., 2013) provide slightly different cosmological parameters and, in particular, a larger value for Ω_m and a lower one for σ_8. As shown in previous work (Wang, 2008; Guo et al., 2013), however, these differences are expected to have little influence on model predictions, once model parameters are tuned to reproduce a given set of observables in the local Universe.

The particle mass is $m_{DM} = 8.61 \times 10^8 M_\odot h^{-1}$ for MSI and $m_{DM} = 6.89 \times 10^6 M_\odot h^{-1}$ for MSII, and the box size length $L_{box} = 500\mathrm{c}\,\mathrm{Mpc}\,\mathrm{h}^{-1}$ and $L = 100\mathrm{c}\,\mathrm{Mpc}\,\mathrm{h}^{-1}$, respectively. In Fig. 3.1, we show the halo mass function (HMF) predicted from the two simulations at $z = 2$, where the halo mass is defined as the mass contained in a sphere which encloses an overdensity corresponding to 200 times the critical density of the Universe (M_{200}). In the following, we will consider as resolved all haloes that contain at least 150 particles. This corresponds to $\sim 10^{11} M_\odot/h$ for the MSI and $\sim 10^9 M_\odot/h$ for the MSII. Below, we will combine the two simulations by selecting galaxies in haloes more massive than $10^{11.5} M_\odot$ from the MSI, and those residing in less massive haloes from the MSII. To investigate equal physical volumes in the MSI and MSII, we will subdivide the MSI box in 125 subboxes, with volume equal to that of the MSII.

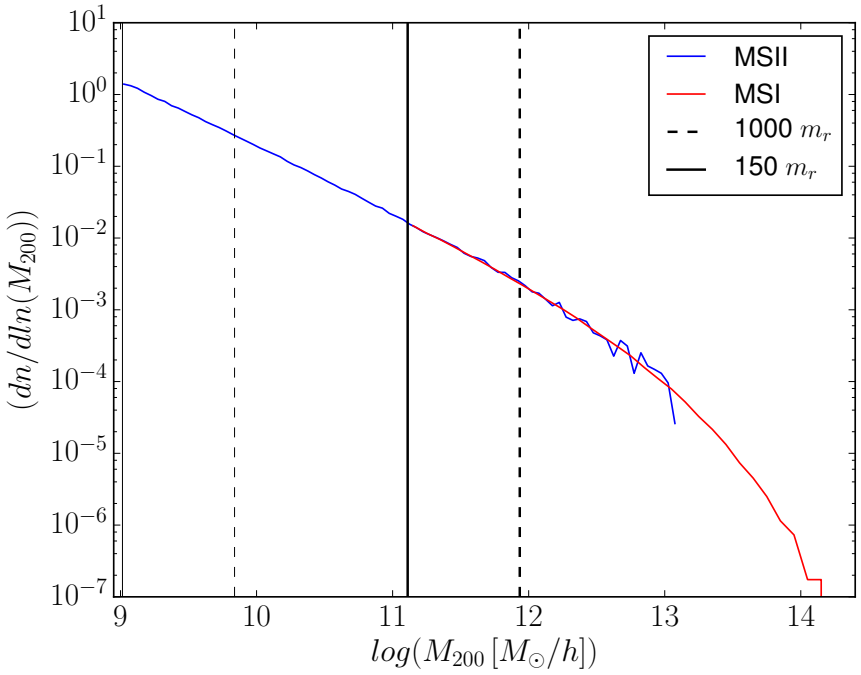

Fig. 3.1: Comparison between the halo mass function, at $z = 2$, derived from the MSI and MSII (red and blue solid lines, respectively). The vertical solid and dashed lines correspond to 150 times and 1000 times, respectively, the DM particle mass for the two simulations. In our analysis, we will assume that haloes with more than 150 particles are well resolved in both simulations.

3.2.2 The semi-analytic model GAEA

The GAlaxy Evolution and Assembly (GAEA) semi-analytic model, at the basis of this work, is an evolution of the model originally described in De Lucia & Blaizot (2007), with significant updates that have been published in the last years (see, in particular, De Lucia et al., 2014; Hirschmann et al., 2016). In this study, we use the version of the model that includes an explicit treatment of the partition of cold gas in its atomic and molecular components (Xie et al., 2017). Specifically, we adopt the fiducial run presented in the work by Xie et al. (2017), based on the empirical prescriptions by Blitz & Rosolowsky (2006). We will refer to this as the BR run in the following.

The GAEA model describes the evolution of four different baryonic reservoirs associated with a dark matter halo: (i) a hot gas reservoir that can grow due to cosmological accretion and stellar feedback, and from which gas cools

onto the gaseous disks of central galaxies; (ii) a cold gas component associated with model galaxies from which stars form, and whose mass is affected by gas recycling due to stellar evolution and by stellar feedback; (iii) a stellar component for each model galaxy; and (iv) an ejected component that stores the gas that has been removed from the inter-stellar medium (ISM) of galaxies (i.e. cannot participate to star formation), and that can be later re-accreted onto the hot component associated with the parent dark matter halo.

The BR prescription, described by Xie et al. (2017), allows a partition of the cold gas into atomic (HI) and molecular (H_2) hydrogen, and has been tuned to reproduce the observed HI mass function at z=0. The ratio of molecular to atomic hydrogen, $R_{\rm mol} = \Sigma_{H2}/\Sigma_{HI}$, depends on 4 physical properties of model galaxies: the mass of the *cold gas* ($M_{\rm CG}$, that in our model corresponds to gas with temperature below 10^4 K), the galaxy stellar mass ($M_\star$), the size of the gaseous disc ($R_{\rm CG,d}$), and the size of the stellar disc ($R_{\star,d}$). Using the empirical relation by Blitz & Rosolowsky (2006), the molecular fraction can be expressed as:

$$R_{mol} = \left(\frac{P_{ext}}{P_0}\right)^{\alpha}$$

where P_0 is the external pressure of molecular clumps and its logarithmic value is assumed to be $log(P_0/k_B[\mathrm{cm}^{-3}\mathrm{K}]) = 4.54$,

$\alpha = 0.92$, $P_{\rm ext} = \frac{\pi}{2} G\Sigma_{\rm CG}\left[\Sigma_{\rm CG} + f_\sigma \Sigma_\star\right]$, $\Sigma_\star$ is the stellar surface density, and Σ_{CG} is the cold gas surface density. The latter is estimated in 21 logarithmic annuli (see original paper by Xie et al. 2017 for details).

In our model, $R_{\rm CG,d}$ and $R_{\star,d}$ are estimated from the specific angular momentum of the gaseous ($J_{\rm CG}$) and stellar ($J_\star$) disk component, respectively, assuming both are well described by an exponential profile:

$$R_{\rm CG,d} = \frac{J_{\rm CG}/M_{\rm CG}}{2V_{max}} \tag{3.1}$$

$$R_{\star,d} = \frac{J_\star/M_\star}{2V_{max}} \tag{3.2}$$

where V_{max} is the maximum circular velocity of the dark matter halo.

Fig. 3.2 shows the HI mass function predicted by GAEA at $z = 0$, and compares model predictions with observational results by Zwaan et al. (2005)

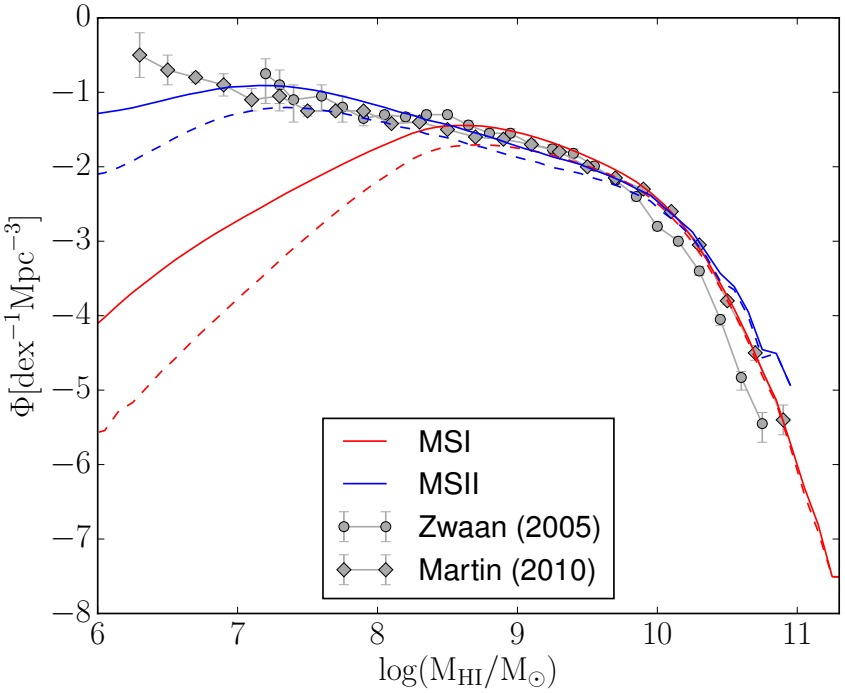

Fig. 3.2: The HI mass function predicted for the MSI (red) and MSII (blue) simulations at $z = 0$. Solid and dashed lines are used for all model galaxies and for centrals only, respectively. Dark grey symbols with error bars show the observational measurements by Zwaan et al. (2005) and Martin et al. (2010). These are based on the blind HI surveys HIPASS (Meyer et al. 2004, limited to $z < 0.04$) and ALFALFA (Giovanelli et al. 2005, limited to $z < 0.06$), respectively. We apply the same stellar mass cuts adopted in Spinelli et al. (2020)

and Martin et al. (2010). The model runs used in this chapter are based on the Millennium I (red line) and Millennium II (blue line) simulations (see next section), that resolve DM haloes down to $\sim 10^{11} M_\odot$ and $\sim 10^9 M_\odot$ respectively. We consider as completeness limit for the cold gas mass at $z = 0$ the values $M_{\rm CG} \sim 10^8 M_\odot$ and $M_{\rm CG} \sim 10^7 M_\odot$ for the MSI and MSII, respectively (for details, see Spinelli et al. 2020).

As mentioned above, the observed HI mass function in the local Universe has been used as the primary constraint for the BR model. Previous works have shown that the same model is able to reproduce a number of important additional observational constrains including scaling relations between the atomic/molecular mass and stellar mass, and the observed evolution of the

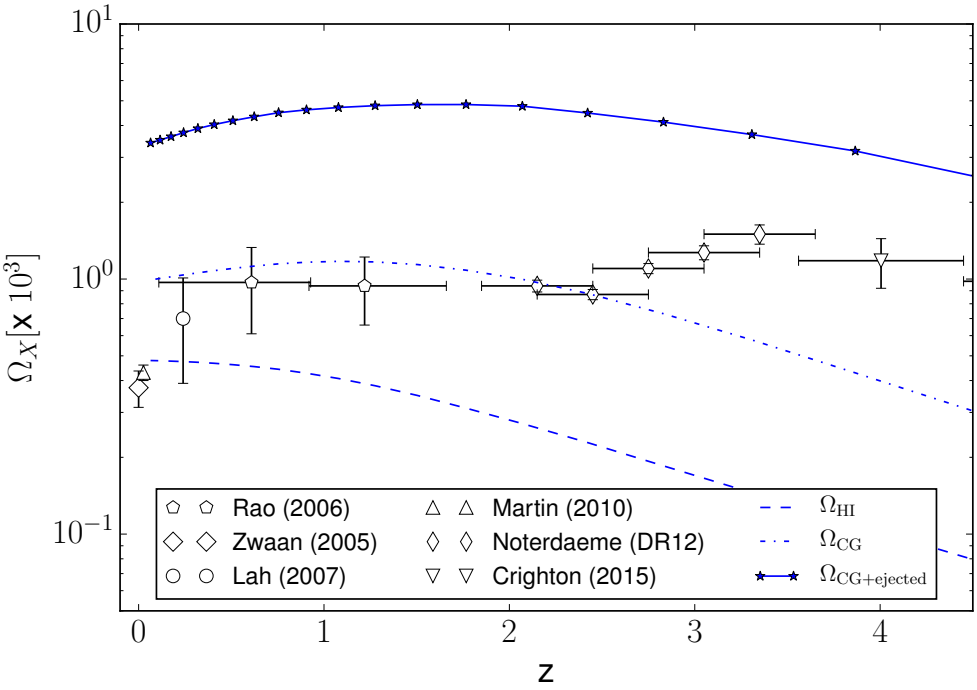

Fig. 3.3: The comoving density evolution of the atomic hydrogen, cold and ejected gas (dashed, dot-dashed, and solid lines, respectively), obtained by summing the corresponding components of all model galaxies down to the resolution limits of the two simulations (see text for details). Model predictions are compared with observational measurements of $\Omega_{\rm HI}$ collected by Crighton et al. (2015).

mass-metallicity relation up to $z \sim 3$ (Hirschmann et al., 2016; Xie et al., 2017; Zoldan et al., 2017). This is relevant for our study that will include an analysis of the metallicities predicted for DLAs.

3.2.3 HI cosmic density

We have estimated the comoving density of the atomic hydrogen, and cold and ejected gas ($\Omega_{\rm HI}$, Ω_{CG} and $\Omega_{ejected}$) in our simulated Universe, summing the corresponding gaseous components of all model galaxies residing in haloes above our adopted resolution limits (for each observable component X, $\Omega_{\rm X}(z) = \frac{\rho_x\,(z)}{\rho_c(0)}$). In particular, we have summed the comoving gas density measured in the MSII box, considering haloes in the mass range $10^{9.2} \leq M_{200} < 10^{11.5} M_{\odot}$ to the average comoving gas density measured

from haloes with $M_{200} \geq 10^{11.5} M_{\odot}$ in the 125 sub-boxes of the MSI, each with a volume equal to the volume of the MSII box. Fig. 3.3 shows these model predictions together with observational measurements of $\Omega_{\rm HI}$ from Crighton et al. (2015). In this figure, we have corrected for the critical density value corresponding to the cosmology adopted by Crighton et al. (2015).

Our simulated estimate of Ω_{HI} is a factor ~ 2.5 below the observational estimates based on DLA surveys up to $z \sim 2$, and further decreases at higher redshift. The low z behaviour of the predicted Ω_{HI} is not surprising, because the GAEA model is tuned to reproduce the HI mass function observed in the local Universe by Martin et al. (2010), whose integrated value is a factor ~ 2 lower than the HI cosmic density estimate by Lah et al. (2007) at $z \sim 0.24$ (also based on emission lines measurements), and than estimates based on statistical analysis of DLAs at higher redshift (Rao et al., 2006; Noterdaeme et al., 2012).

The decrease of Ω_{HI} at high redshift ($z > 3$) is more difficult to explain. Spinelli et al. (2020) show that the largest contribution to Ω_{HI} in our model is given by haloes with mass $10^{10} M_{\odot} \leq M_{200} \leq 10^{12} M_{\odot}$, and that Ω_{HI} decreases with increasing redshift for more massive haloes while it flattens for less massive haloes. The decrease of Ω_{HI} at higher redshift is found also for independent semi-analytical models that consider a similar mass range of dark matter haloes contributing to the HI density (e.g. Lagos et al., 2011; Berry et al., 2014).

A possible solution to this problem is to increase the contribution of intermediate and low-mass haloes to Ω_{HI} at high redshift, that we may be underestimating because of the adopted physical prescriptions and resolution limits of our simulations. It is difficult to quantify precisely the impact of resolution on our results, as it can affect both the missing HI content of the unresolved isolated haloes and the HI content of the satellite galaxies hosted in the resolved haloes. Considering the resolution limit of the MSII, and the observed scaling relation between the HI to stellar mass ratio and galaxy stellar mass, we expect that the largest contribution should come from the HI content of unresolved haloes. In the next section, we explain how we compute an estimate of such a contribution.

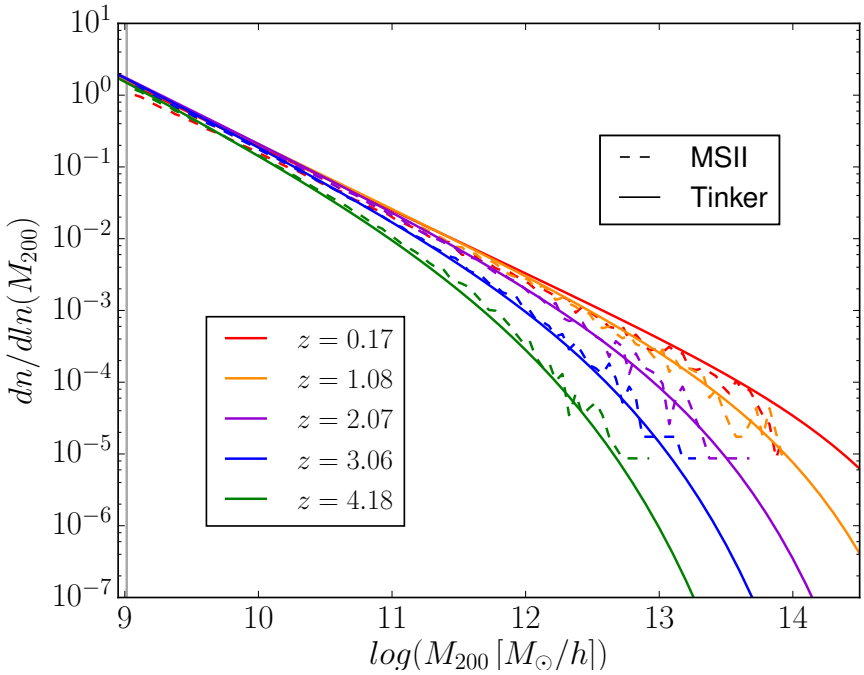

Fig. 3.4: Comparison between the Tinker halo mass function (solid lines) and that estimated from the MSII (dashed lines), at 5 different redshifts, listed in the legend.The vertical line marks the resolution limit of the MSII simulation.

3.2.4 Minimal HOD model

To quantify whether low-mass haloes (i.e. below $\sim 10^9 M_\odot$) can significantly increase the HI density in our simulated Universe, we populated the MSII box with haloes below its resolution using a simple halo occupation distribution (HOD model - see Berlind & Weinberg 2002 for an historical review).

The number of low-mass haloes to be added, and their mass distribution, have been derived integrating the HMF by Tinker et al. (2008) in the range $10^8 M_\odot \leq M_{200} < 10^{9.2} M_\odot$, and using the cosmological parameters adopted for the Millennium simulations. We have checked that the shape and normalization of the Tinker HMF are consistent with those derived from the MSII and MSI. This can be appreciated in Fig. 3.4, where we compare the Tinker HMF (solid lines) with that measured from the MSII (dashed lines), at 5 different redshifts.

Considering the low mass of the haloes treated with the HOD model, we

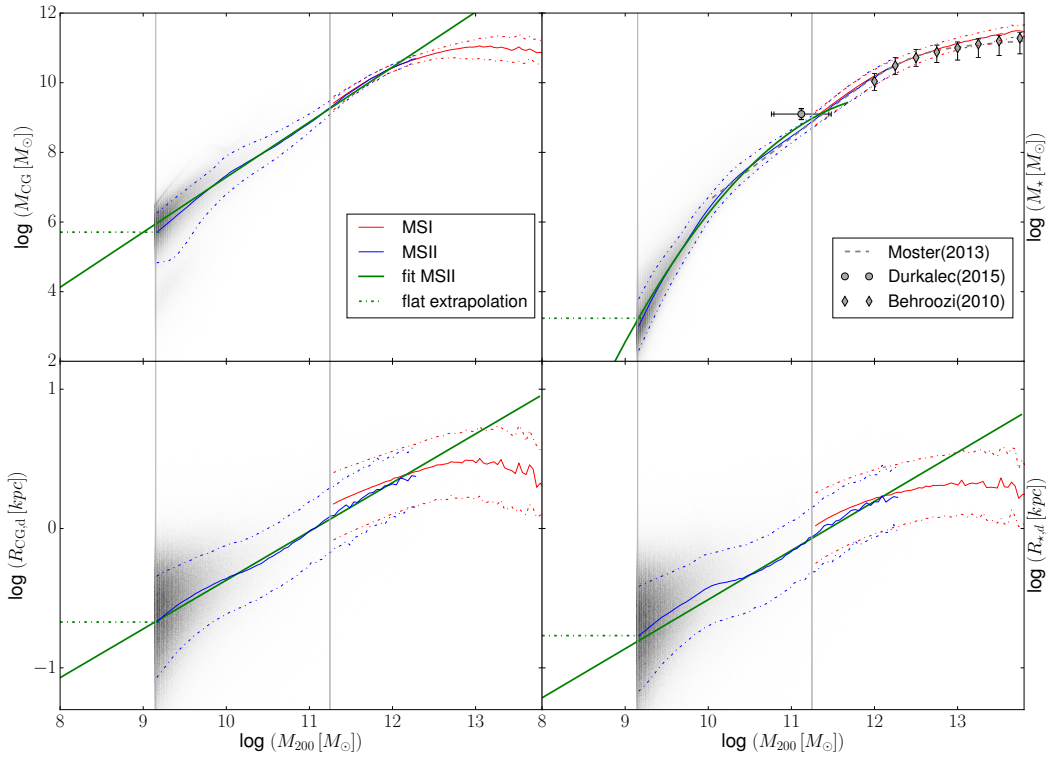

Fig. 3.5: Scaling relation for $M_{\rm CG}$, $M_\star$, $R_{{\rm CG},d}$, and $R_{\star,d}$ as a function of halo mass for central galaxies in MSI (red) and MSII (blue), at redshift $z = 2.07$. The grey color gradient highlights the number density of MSII central galaxies. The solid and dot-dashed lines show respectively the median and percentiles (16th and 84th) of the distributions. The solid green lines show the linear fit for all relations extracted from the MSII, except for the $M_\star$ vs M_{200} relation. In this case, we use a polynomial fit of second order. The dot-dashed lines are flat extrapolations of the scaling relations, normalized to median values corresponding to the lowest halo mass bin resolved in the MSII. In each panel, the vertical solid thick (thin) line shows the resolution limit of the MSI (MSII). For the SMHM relation we show also observational estimates (Behroozi et al., 2010; Durkalec et al., 2015) and the fitting function derived by Moster et al. (2013) for central galaxies.

have populated them only with central galaxies, since we do not expect that they host satellites, and distributed them at random positions inside the MSII box. We assign 5 physical quantities to galaxies in the HOD model: stellar mass($M_\star$), cold gas mass ($M_{\rm CG}$), scale radius of the gaseous disk ($R_{{\rm CG},d}$),

scale radius of the stellar disk ($R_{\star,d}$), and abundance ratio $[\mathrm{Fe/H}]$. These quantities are derived extrapolating the scaling relations obtained from our semi-analytic model run on the MSI and MSII. The scaling relations for the first four quantities are shown in Fig. 3.5 for $z = 2$ (these scaling relations evolve slowly as a function of redshift), while the extrapolation of $[\mathrm{Fe/H}]$ is treated in detail in subsection 3.3.2. In the mass regime where MSI and MSII overlap we observe a nice convergence of the scaling relations (e.g. the difference between galaxy stellar mass of MSI and MSII is less than 10 $\%$ for haloes with $M_{200} \sim 10^{11} M_{\odot}$). Fig. 3.5 also shows that the predicted SMHM relation is in good agreement with the observational estimates (Behroozi et al., 2010; Moster et al., 2013; Durkalec et al., 2015).

We have considered two different extrapolations of the predicted scaling relations at each of the snapshots analysed: (i) a linear fit of the median relation obtained for galaxies in MSII (a second order polynomial for the stellar mass - halo mass relation); (ii) a flat extrapolation normalized to the value obtained for the smallest haloes in the MSII.

Based on the values extrapolated for $M_{\star}$, M_{CG}, $R_{\star,d}$, and $R_{\mathrm{CG},d}$, we then estimate the molecular fraction using the empirical relation by Blitz & Rosolowsky (2006) in 21 annuli. For each galaxy in the HOD catalogue, we store the integrated molecular gas fraction in the disk (R_{mol}), that we use to estimate the atomic gas mass. Fig. 3.6 shows the evolution as a function of redshift of the comoving density of HI and cold gas (solid and dashed lines, respectively) computed considering all DM haloes from the MSI, MSI and HOD model. The contribution to Ω_{CG} coming from the HOD galaxies becomes non-negligible only at relatively high redshifts (e.g. for $z > 4.5$), and only when considering a flat extrapolation of the scaling relations. The contribution to the cosmic density of neutral hydrogen is dominated by MSI haloes up to redshift $z \simeq 2.3$, when the MSII starts dominating. For the cold gas, the cross-over between the MSI and MSII takes place at $z \gtrsim 3.5$.

We have studied the effect of different halo mass cuts on the cosmic HI content in our simulated Universe, finding little differences. In the following, we adopt the following fiducial cuts: we select haloes with $\log(\frac{M_{200}}{M_{\odot}})$ in the range $[8, 9.2)$ from the HOD, haloes with $\log(\frac{M_{200}}{M_{\odot}})$ in the range $[11.5, max)$ from MSI, and haloes from the MSII in the intermediate regime.

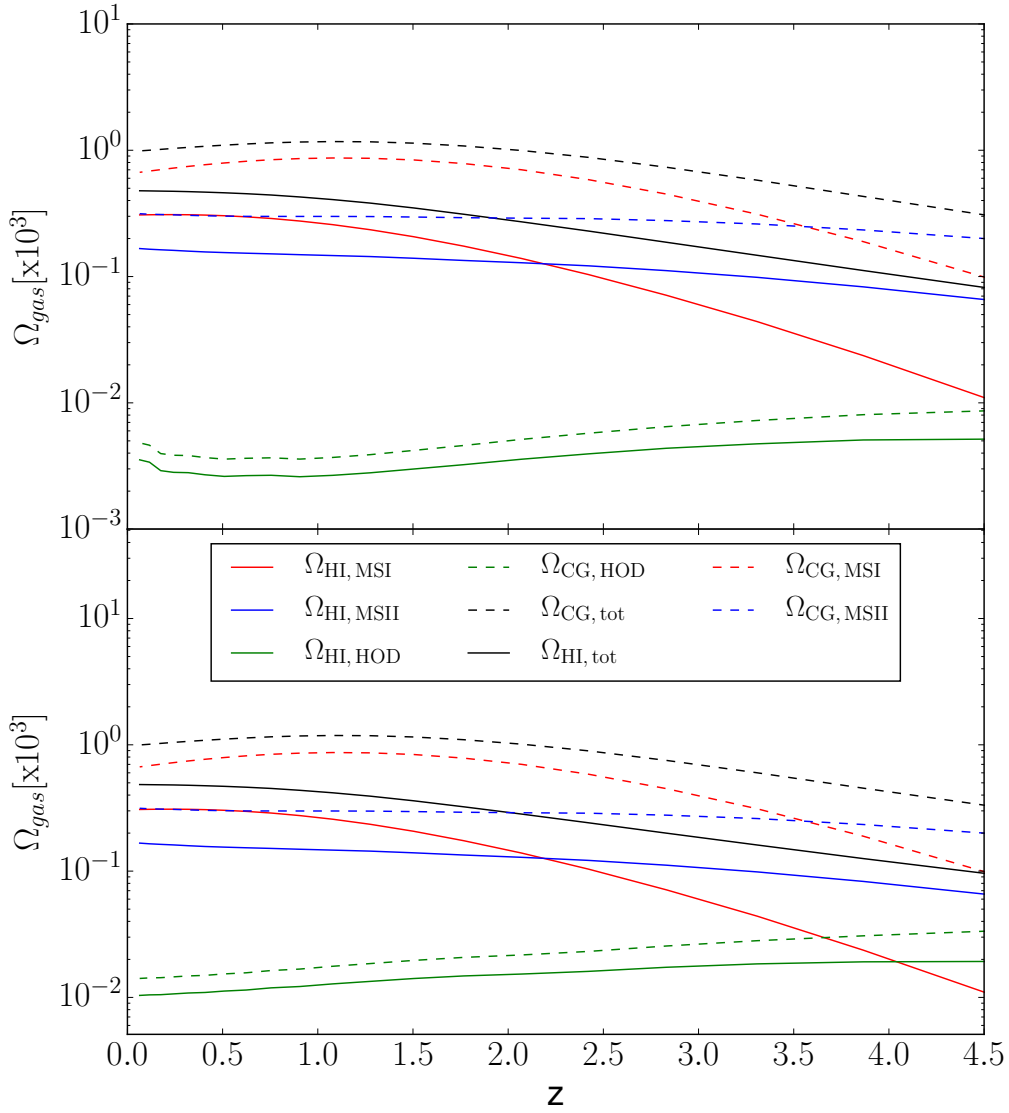

Fig. 3.6: Evolution of the comoving density of atomic hydrogen and cold gas content of model galaxies (HI solid, CG dashed), residing in massive haloes of MSI (red), intermediate-mass haloes of MSII (blue) and low-mass haloes of our HOD extension (green). The black solid (dashed) line shows the total content of HI (CG) in our simulated universe. The top panel shows the contribution coming from HOD galaxies when considering the linear (or 2nd-order for the stellar mass - halo mass relation) extrapolation of the scaling relations obtained from the MSII galaxies, while the bottom panel corresponds to the flat extrapolation of the scaling relations.

3.3 Creation of the simulated DLA catalog

In order to produce samples of simulated DLAs to be compared with observational data we have thrown random lines of sight (LOS) in the volume of the composite simulation described in the previous section. To cover a large halo mass range $(10^8 - 10^{15} M_\odot)$, we consider together the galaxies hosted in the DM haloes selected from the MSI, MSII and in those added using the HOD, according to the halo mass cuts described in the previous section. The physical properties assigned to HOD galaxies are derived adopting the flat extrapolation.

We have subdivided the volume of MSI in 125 sub-boxes, of volume equal to that of the MSII box ($L_{box} = 100 h^{-1}\mathrm{Mpc}$), and constructed at each redshift analysed 125 realizations that differ only for the MSI contribution. In this way, it is possible to investigate the impact of the cosmic variance on the DLA observables considered in this study, for the DM haloes that are well resolved in the MSI (i.e. with $M_{200} > 10^{11} M_\odot$). Since cosmic variance is more important for rarer (i.e. more massive) systems, we expect that it does not play an important role for the intermediate mass haloes that are selected from the MSII simulation.

For each simulation snapshot in the redshift range of interest, we throw 100,000 random LOS, parallel to the $z-$axis, for each of the 125 realizations considered. This provides us with 125 simulated DLA catalogs.

3.3.1 $\mathrm{N_{HI}}$ estimate

For each galactic disk, we assume that the gas density profile follows a double-exponential profile[1]:

$$\rho_{\mathrm{CG}}(r, z) = \rho_0 \; e^{-r/R_{\mathrm{CG},d}} \; e^{-z/z_0} \tag{3.3}$$

where ρ_0 is the normalization of the 3D density profile for the gaseous disk, $R_{\mathrm{CG},d}$ and z_0 are the scale-radius and the scale-height of the gaseous disk, respectively.

For the scale-height parameter z_0, we apply a linear dependence on the

[1]We have tested that assuming an isothermal vertical profile for the gas in the ISM (van der Kruit & Freeman, 2011), or alternative vertical profiles suggested in the literature (see Appendix B), does not affect significantly our results.

scale radius:

$$z_0 = \frac{R_{\rm CG,d}}{A}, \tag{3.4}$$

and test two different values of the fudge factor A: $= 7.3$ and $= 4$. The former choice relies on observational relations valid for stellar disks in the local Universe (Kregel et al., 2002), while the latter is motivated by observations of thicker galactic stellar disks at $z \sim 2$ (Elmegreen et al., 2017).

The HI density profile can be written as:

$$\rho_{\rm HI}(r,z) = (1 - f_{mol}(r)\,)\,\rho_{\rm CG}(r,z) \tag{3.5}$$

where the molecular fraction f_{mol} has been estimated using the BR prescription (described in Section 3.2.2), in 21 logarithmic radial bins between $r = 0$ and $r = 10\,R_{\rm CG,d}$. When a given LOS intersects a galaxy with a distance (impact parameter) $b \leq 10 R_{\rm CG,d}$, the hydrogen column density ($\rm N_{HI}$) contributed by the galaxy can be estimated by integrating the HI density profile along the LOS. The value of $\rm N_{HI}$ depends then on the impact parameter and on the inclination of the galactic plane with respect to the LOS.

The assumption that cold gas in model disk galaxies is distributed according to an exponential density profile is in good agreement with observational findings (e.g. Wang et al., 2014).

We have also considered the contributions from close galaxies/pairs to each absorption feature. Adopting a FoF-like merging algorithm, we summed all column densities of absorbing systems, intersected by the same LOS, with a maximum velocity offset of $\Delta v \leq 2000\ km/s$. Our merging algorithm works as follows: we firstly subdivide the systems along the same LOS into groups of close systems sorted along the z-coordinate. Then, we merge the two nearest systems in each group, estimate the barycentre of the pair and re-estimated the distance between the first merged system and the other systems in each group. If necessary, we repeat the merging process and re-iterate until there is no other pair to merge.

The estimated fraction of DLAs originated from multiple systems is large (more than 70% at $z = 2$ for the $2M - 2R$ model and more than 50% for the fiducial model). However, in most cases one single galaxy contributes significantly more than the others. In particular, if we consider only systems with column density $N_{\rm HI} \geq 10^{17}\,atoms\,{\rm cm}^{-2}$, in 87% (84%) of the cases more than 80% of the total hydrogen column density comes from one single

galaxy while the cases where the contribution of each single galaxy is less than 50% represent only 1% (0.5%) of the all cases for the $2M-2R$ model (for the fiducial model). Therefore the distribution of simulated DLA column densities is not significantly affected by the blending of close absorption features.

3.3.2 Assigning metallicity to DLAs

The GAEA model adopts a detailed chemical enrichment scheme that accounts for the finite lifetime of stars and the non-instantaneous recycling of metals, gas, and energy (De Lucia et al., 2014).

As discussed in previous work, the fiducial model used here is able to reproduce the observed evolution of the correlation between galaxy stellar mass and cold gas metallicity, up to $z \sim 2$ (Hirschmann et al., 2016; Xie et al., 2017). This is an important achievement, met by only a few recently published theoretical models (see discussion in Somerville & Davé, 2015). Our study offers an additional test to the model.

As commonly done in DLA studies, we use the iron over hydrogen abundance ratio, $\left[\frac{\mathrm{Fe}}{\mathrm{H}}\right]$, as a proxy for the metallicity of the gaseous disks of our simulated galaxies. GAEA assumes a uniform distribution of the metals in the different baryonic components. So we can write:

$$\left[\frac{\mathrm{Fe}}{\mathrm{H}}\right] = \log\left(\frac{\mathrm{M_{Fe,d}}\,\mu_{\mathrm{H}}}{\mathrm{M_{HI,d}}\,\mu_{\mathrm{Fe}}}\right) - \log\left(\frac{\mathrm{M_{Fe}}\,\mu_{\mathrm{H}}}{\mathrm{M_{HI}}\,\mu_{\mathrm{Fe}}}\right)_{\odot} \tag{3.6}$$

where $\mathrm{M_{Fe,d}}$ and $\mathrm{M_{HI,d}}$ are the masses of Fe and HI in the cold gaseous disk of each galaxy, while μ_{Fe} and μ_{HI} are the corresponding mean atomic weights. $[\mathrm{Fe/H}]_{\odot}$ is the solar abundance ratio, that we take from Asplund et al. (2009).

Fig. 3.7 shows the relation between the abundance ratio $[\mathrm{Fe/H}]$ and $log(\mathrm{M_{200}})$, for the central galaxies in the MSI and MSII. There is a good convergence between MSI and MSII in the galaxy mass range $10^{8.5} < M_{\star} < 10^{10} M_{\odot}$. For the extrapolation of the $[\mathrm{Fe/H}] - \mathrm{M_{200}}$ relation to galaxies inside the haloes sampled by the HOD, we have used a linear regression in the mass range $10^{9.2} < \mathrm{M_{200}} < 10^{10.6} [\mathrm{M_{\odot}}]$, i.e. after the step-like feature visible in the figure. This feature arises mainly as a consequence of a specific assumption of our galaxy formation model: for $\mathrm{M_{200}} \leq 5 \times 10^{10} \mathrm{M_{\odot}}$, 95% of the new metals

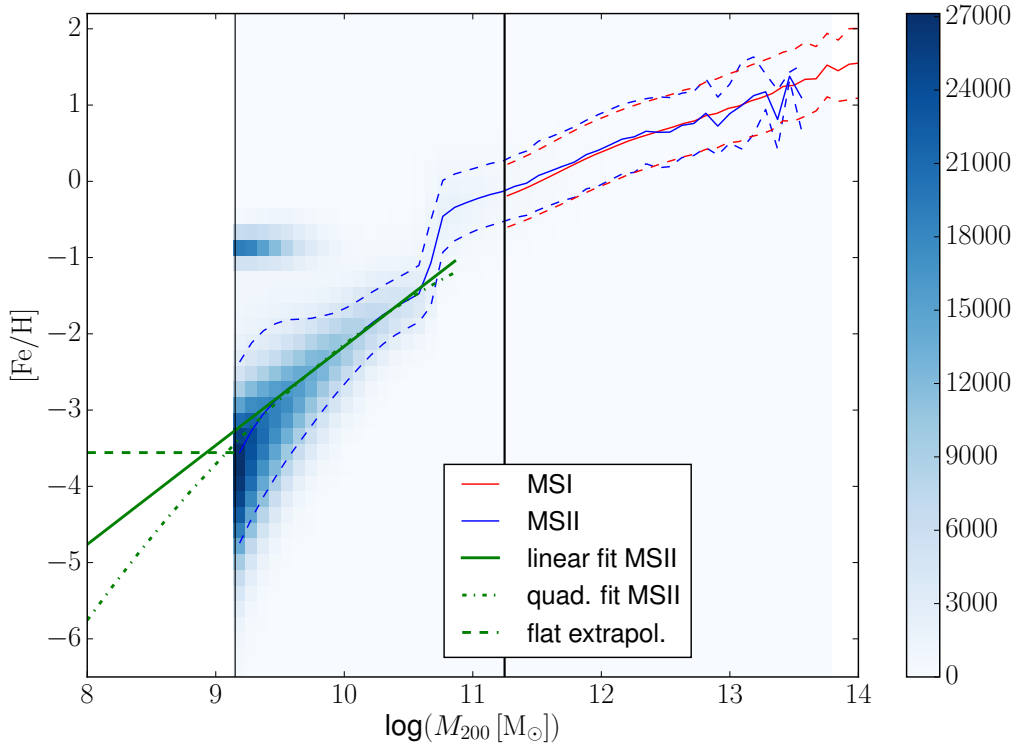

Fig. 3.7: [Fe/H] as function of $log(M_{200})$ for central galaxies, at $z = 2$. The solid blue (red) line shows the mean relation for MSII (MSI) galaxies. The green solid, dashed and dot-dashed lines show the extrapolated linear fit to the mean relation measured for the MSII, a flat extrapolation, and an extrapolation based on a quadratic fit to the MSII results. The color coding quantifies the number density of the MSII central galaxies. The vertical lines show the resolution limits of the MSI and MSII.

are ejected directly into the hot phase, instead of being mixed with the cold-gas in the ISM, as assumed for more massive haloes. This assumption was motivated by results from hydrodynamical simulations (Mac Low & Ferrara, 1999) and was helpful, in previous versions of our models, to reproduce the metal content of satellites in Milky-Way-like haloes (Li et al., 2010).

Considering the large $1 - \sigma$ scatter of the predicted relation, we have applied it to the extrapolated values of the abundance ratio for HOD galaxies. Once assigned the atomic hydrogen mass to HOD galaxies (following the procedure described in the previous section), we can use the extrapolated

abundance ratio to assign an iron mass to each HOD galaxy.

We have also considered the effect due to the presence of a metallicity radial gradient. Specifically, we have assumed a slope consistent with the observational study by Christensen et al. (2014b):

$$\Gamma = -0.022 \text{ dex kpc}^{-1} \tag{3.7}$$

Christensen et al. (2014b) and Rhodin et al. (2018) found an almost universal metallicity gradient for a sample of DLAs observed in the redshift range $(2-3.5)$. Stott et al. (2014) found at $z \sim 1$ a slight correlation between the metallicity gradient and the sSFR (but see Carton et al., 2018; Ma et al., 2017, for a different view). Here we test if we are able to recover the observed trends using the simplest assumption of a universal metallicity gradient.

The iron over hydrogen abundance ratio can be estimated for each DLA, for a given impact parameter $b = r$, applying the following formula:

$$\left[\frac{\text{Fe}}{\text{H}}\right](r) = log_{10}(Z_{\text{DLA},0}) - log_{10}(Z_{\odot}) - \Gamma r \tag{3.8}$$

where

$$Z_{\text{DLA},0} = Z_{\text{DLA}}(r=0) =$$

$$= \frac{\mu_{\text{HI}}}{\mu_{\text{Fe}}} \frac{\text{M}_{\text{Fe}}(1-11\,e^{-10})}{\text{M}_{\text{HI}}} \frac{(1-<f_{mol}>)}{\int_0^{10} dy(1-f_{mol}(y))} e^{-(1+ln(10)\Gamma r_s)y}$$

and $y = r/r_s$, while $r_s = R_{\text{CG},d}$ and f_{mol} is the molecular fraction.

3.4 Simulated DLA properties

In this section, we compare the properties of the DLAs in our simulated Universe with those estimated from observational data. For each property derived in this section we have combined the MSI and MSII simulation as explained in Section 3.3.

3.4.1 The column density distribution function

The column density distribution function (CDDF) is defined as the number of absorbers observed per unit redshift path and column density interval:

$$f(N_{\rm HI}, X) dX dN_{\rm HI} = n_{abs}(N_{\rm HI}, X), \tag{3.9}$$

where the absorbing path dX is defined as $dX = \frac{H_0}{H(z)}(1+z)^2 dz$, in terms of the redshift path dz.

The CDDF plays, in absorption line studies, a similarly central role (and provides a similarly 'vague' information) as the luminosity function in galaxy evolution studies. The analytic model of the CDDF proposed by Schaye (2001) (devised for over-densities that cannot self-shield from the UV background), together with results from cosmological simulations (e.g. Altay et al., 2011), indicate that systems of a given column density originate from dramatically different over-densities. Nevertheless, higher column densities systems are typically connected to denser gas that, in general and average sense, tends to lie closer to galaxies. It has been argued that the steepest part of the CDDF, made of the densest absorbers, may be particularly sensitive to stellar feedback and stellar evolution (Rosenberg & Schneider, 2003; Bird et al., 2014).

Fig. 3.8, in the upper panel, shows the CDDF derived from our simulated absorbers in the redshift range $2 < z < 3$, for our fiducial combination of halo mass cuts (see subsection 3.2.4) applied to fiducial GAEA model. Model predictions are compared with observational estimates by Noterdaeme et al. (2012).

The figure shows a significant discrepancy between our fiducial model and the observed CDDF, in particular below $\log(N_{\rm HI}) < 21$. This discrepancy motivated us to test the dependence of the CDDF on the physical properties of simulated galaxies, and in particular the scale radius of the gaseous disk and the cold gas mass. The lower panel of Fig. 3.8 shows results obtained multiplying by a factor 2 both the scale radius and the cold gas mass of all model galaxies (the 2M-2R model introduced above).Results from this ad-hoc modifications are in very good agreement with observations at $20.0 < \log(N_{\rm HI}) < 22.2$, for the redshift range considered. We have verified that the better agreement with observational data is mainly driven by the increase of scale radius, that leads to a larger galaxy cross-section (i.e. a

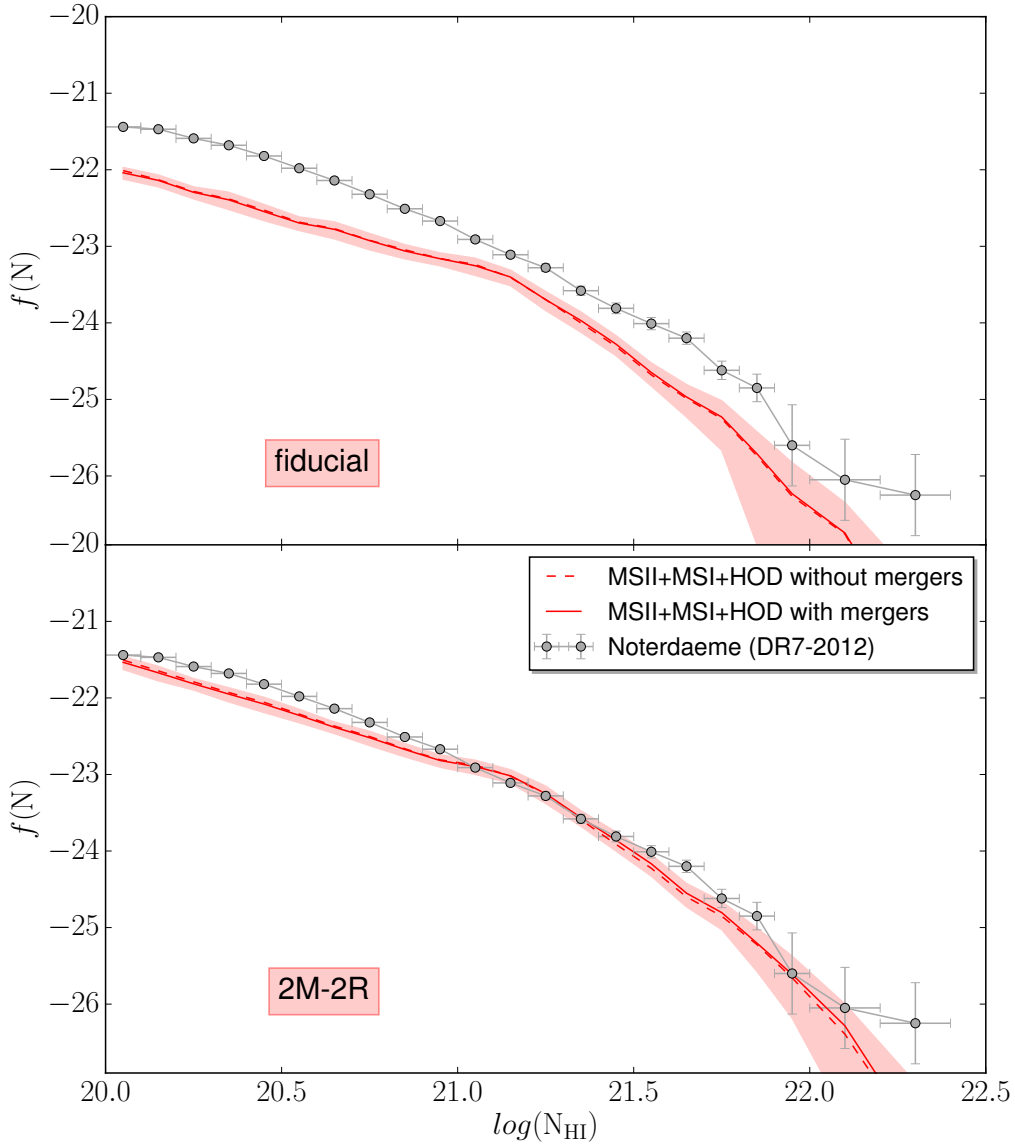

Fig. 3.8: Predicted column density distribution function (CDDF) in the redshift range $2 < z < 3$. The top panel shows results based on our fiducial model, while the bottom panel shows the results of the model where $R_s = 2R_{s,\mathrm{orig}}$ and $M_{CG} = 2M_{CG,\mathrm{orig}}$. We estimate the CDDF for each of the 125 realizations described in Sect. 3.3, converting into a redshift interval (dz) the length of each LOS, that is equal to $L_{box} = 100h^{-1}\mathrm{Mpc}$ comoving at all redshifts. The solid red line indicates the average of the CDDFs obtained for all realizations considered in the redshift range of interest (2izi3), while the shaded area highlight the $1-\sigma$ scatter of the distribution.

larger probability of intersecting model galaxies).

3.4.2 The cosmic hydrogen density associated with DLAs

The cosmic hydrogen density associated with DLAs can be computed as:

$$\Omega_{\rm DLA} = \frac{m_H H_0 \sum_i N_i(HI)}{c\rho_c \Delta X}. \tag{3.10}$$

where ρ_c is the critical density at $z = 0$, $m_{\rm H}$ is the mass of the hydrogen atom, and the sum is carried out over all systems with $\log {\rm N(HI)} > 20.3$, across a total absorption path length ΔX.

Fig. 3.9 compares the redshift evolution of the comoving HI density derived from our simulated DLAs, with the observational estimates from Crighton et al. (2015). To be consistent with the observations, we have corrected the values provided by Eq. 3.10 by a factor 1.2 ($\Omega^{\rm HI}_{\rm DLA} = 1.2 \times \Omega_{\rm DLA}$), that takes into account the contribution to the comoving HI density of absorbers with column density below $N_{HI} = 20.3$ (Crighton et al., 2015). Using this correction, $\Omega^{\rm HI}_{\rm DLA}$ turns out to agree remarkably well with $\Omega^{\rm HI}_{\rm gals}$ that is derived summing the HI contribution of all model galaxies. This non trivial result indicates that our model predicts the correct shape for the CDDF (both for the fiducial and the $2M - 2R$ model).

$\Omega^{\rm HI}_{\rm DLA}$ derived from our simulations ('fiducial' model) is, on average, a factor ~ 2.5 below the observational estimates in the redshift range $0 < z < 2$, and it further decreases to about an order of magnitude below the data at $z = 4$. As discussed in Sect. 3.2, the difference is in part due to the fact that our model is tuned to reproduce the HIMF measured in the local Universe (Martin et al., 2010), that gives an estimate of the integrated comoving HI density a factor ~ 2 lower than that derived from DLA observations. The 2M-2R assumption alleviates the discrepancy, at least up to redshift $z \sim 3$, as shown in the lower panel of Fig. 3.9.

At higher redshift, also the predictions from the $2M - 2R$ model exhibit a significant decline, while observations measure little evolution of $\Omega^{\rm HI}_{\rm DLA}$ up to $z \sim 5$. This can be due to different reasons: one hypothesis is that the uniform redistribution of the missing hydrogen, applied in the $2M - 2R$ model, is limited, since it gives too much gas to the more massive haloes, which already reproduce the observations, and too less to the intermediate/ low mass ones. The other possibility is that the contribution of outflows

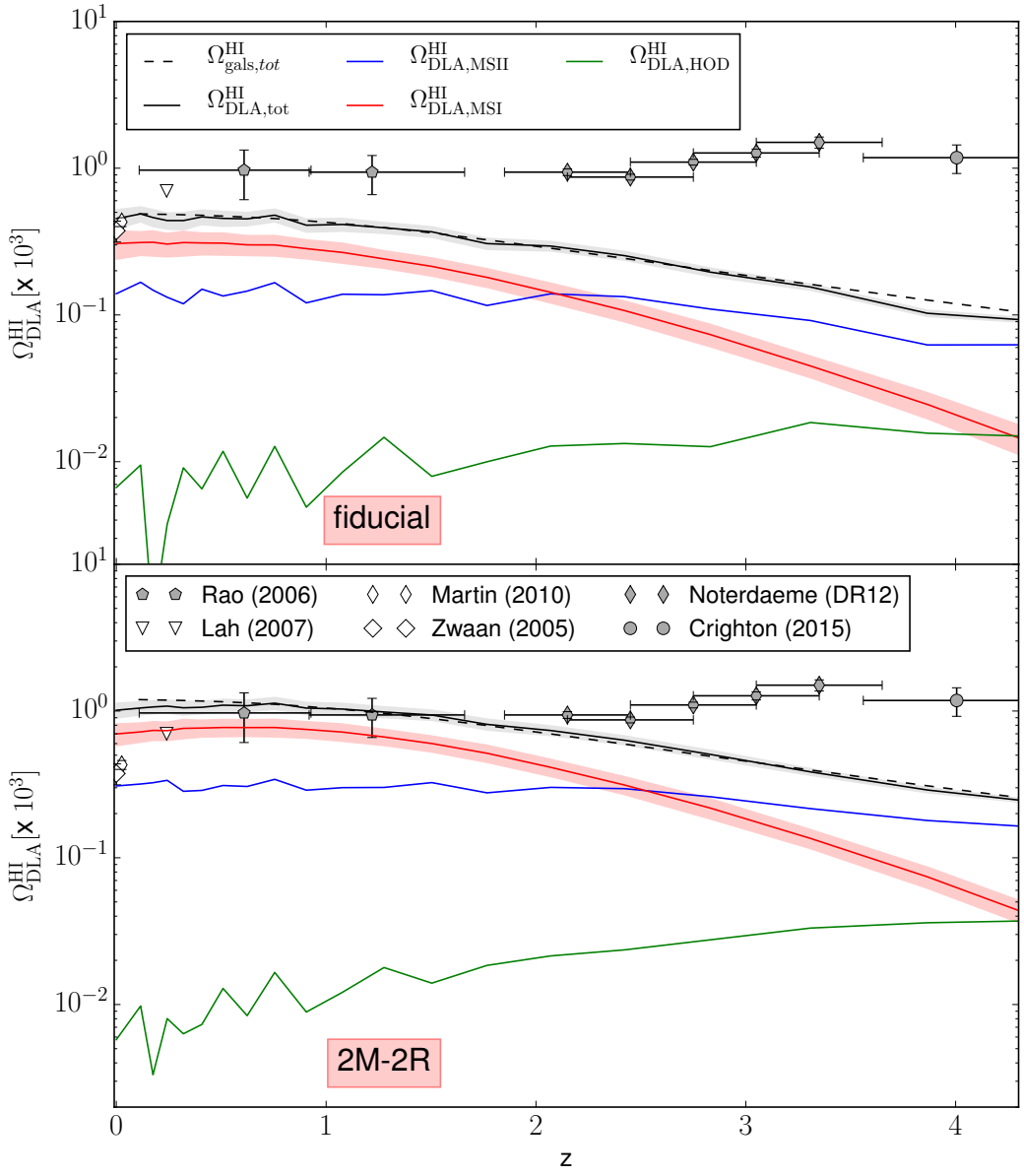

Fig. 3.9: The top (bottom) panel shows the evolution with redshift of $\Omega^{\rm HI}_{\rm DLA}$, in our fiducial (2R-2M) model. We define $\Omega^{\rm HI}_{\rm DLA} = 1.2 \times \Omega_{\rm DLA}$, taking into account the contribution to the comoving HI density of systems with column density lower than the characteristic one of DLAs (Crighton et al., 2015). The solid black line shows the average $\Omega^{\rm HI}_{\rm DLA}$ evolution considering the contribution of all individual systems, while the dashed black line refers to the comoving HI density ($\Omega_{\rm HI}$) of all the galaxies in the box. The three solid lines in red, blue and green refer to the DLAs in the MSI, MSII and HOD respectively. Symbols with error bars show observational data points, taken from the literature as detailed in the legend, and expressed in the cosmology used by Crighton et al. (2015).

and/or filamentary structure becomes more significant at higher redshift (e.g. van de Voort et al., 2012; Fumagalli et al., 2011).

3.4.3 DLA metallicity

Relation between metallicity and $\mathrm{N_{HI}}$

We compare the metallicity of our simulated DLAs with observations taking advantage of the catalog by De Cia et al. (2018), that provides also dust-corrected abundance ratios. As explained earlier, we adopt the iron over hydrogen abundance ratio ($\mathrm{[Fe/H]}$) as a proxy of the metallicity, and we analyze separately the 125 DLA catalogs built (as described in Section 3.3), in the redshift range $2 < z < 3$.

Fig. 3.10 and Fig. 3.11 show $\mathrm{[Fe/H]}$ as a function of $\mathrm{N_{HI}}$ in the redshift range $2 < z < 3$, with lines of different styles contouring the regions enclosing 68, 95 and 99 per cent of the distribution coming out from the stacking of the 125 DLA catalogues. Green symbols with error bars show observational measurements. Fig. 3.10 shows results obtained assuming a uniform distribution of the metals in the gaseous disk, while Fig. 3.11 shows the distribution obtained assuming a universal metallicity gradient (see Sect. 3.3.2).

The simulated abundance ratios appear in somewhat better agreement with the data when we consider a metallicity gradient, in particular at larger metallicity values. Our model, however, predicts a not negligible number of low abundance ratios ($\mathrm{[Fe/H]} < -3.$) systems that are not observed (Prochaska & Wolfe (2009)).

To make the comparison more quantitative, we carry out a two-dimensional Kolmogorov-Smirnov test to quantify the probability that the simulated and observed distributions are extracted from the same sample (i.e. are consistent).

The estimate of the p-value, namely the probability of obtaining the observed distribution assuming the null hypothesis, is $1.16 \cdot 10^{-5}$ ($1.32 \cdot 10^{-5}$) for the $2M - 2R$ run, with (without) a correction for the metallicity gradient. The corresponding value for the fiducial model is $1.77 \cdot 10^{-7}$ ($5.53 \cdot 10^{-7}$). Therefore the hypotesis that the observed DLA metallicities come from the same parent population of the simulated DLAs is on average rejected for both models considered, independently of the metallicity gradient applied.

The discrepancy between observed and simulated data is mainly driven

by the presence, in our model, of low-metallicity systems that are absent in the observed DLA samples. In addition, our simulated DLAs corresponding to large column densities tend to have an average metallicity that is larger than the observed one. This 'shift' in the average metallicity at higher column densities of simulated systems, with respect to that observed, increases in the 2M-2R model.

The excess of low-metallicty systems in our model suggests that the treatment of the chemical enrichment of low-mass haloes (see Fig. 3.7) may be inadequate and should be revised. The difference in the distributions at larger metallicities is more difficult to explain. It is worth noting that the High A_V Quasar survey (HAQ Fynbo et al., 2013; Krogager et al., 2015; Zafar et al., 2015) and the extended-HAQ (Krogager et al., 2016) have shown that the traditional quasar selection used in SDSS is biased against reddened quasars. In addition, the work by Noterdaeme et al. (2015) showed that DLAs associated with large column-densities and metallicities are typically found to exhibit a more significant reddening of the background quasar. Therefore, it is plausible that the combined effect of dust and large atomic hydrogen densities cause a dust-bias in DLA observations, preferentially excluding from the observations DLAs hosted in massive, metal-rich and dusty galaxies. In the $2M-2R$ model, the average metallicity is slightly larger than for the fiducial model, due to the reassignment 'a posteriori' of the scale radius and the mass, which penalizes the low-mass galaxies at intermediate/high column density (see Sec. 3.4.5).

Cosmic metallicity evolution

As observed by Rafelski et al. (2012), the chemical enrichment of DLAs evolves from about 1 to 10 per cent solar from $z \sim 5$ to today. The work by Rafelski et al. (2012) also revealed a statistically significant decline of the DLA average metallicity with increasing redshift, that can be described as $< \Omega_Z >= (-0.26 \pm 0.07)z - (0.59 \pm 0.18)$. This behaviour was confirmed at $z < 4$ by independent measurements (Kulkarni et al., 2007, 2010).

De Cia et al. (2018) also found a similar decrease with redshift, but with a different normalization at low redshift, based on abundance ratios corrected for dust depletion.

We have investigated the evolution of the DLA metallicity by computing

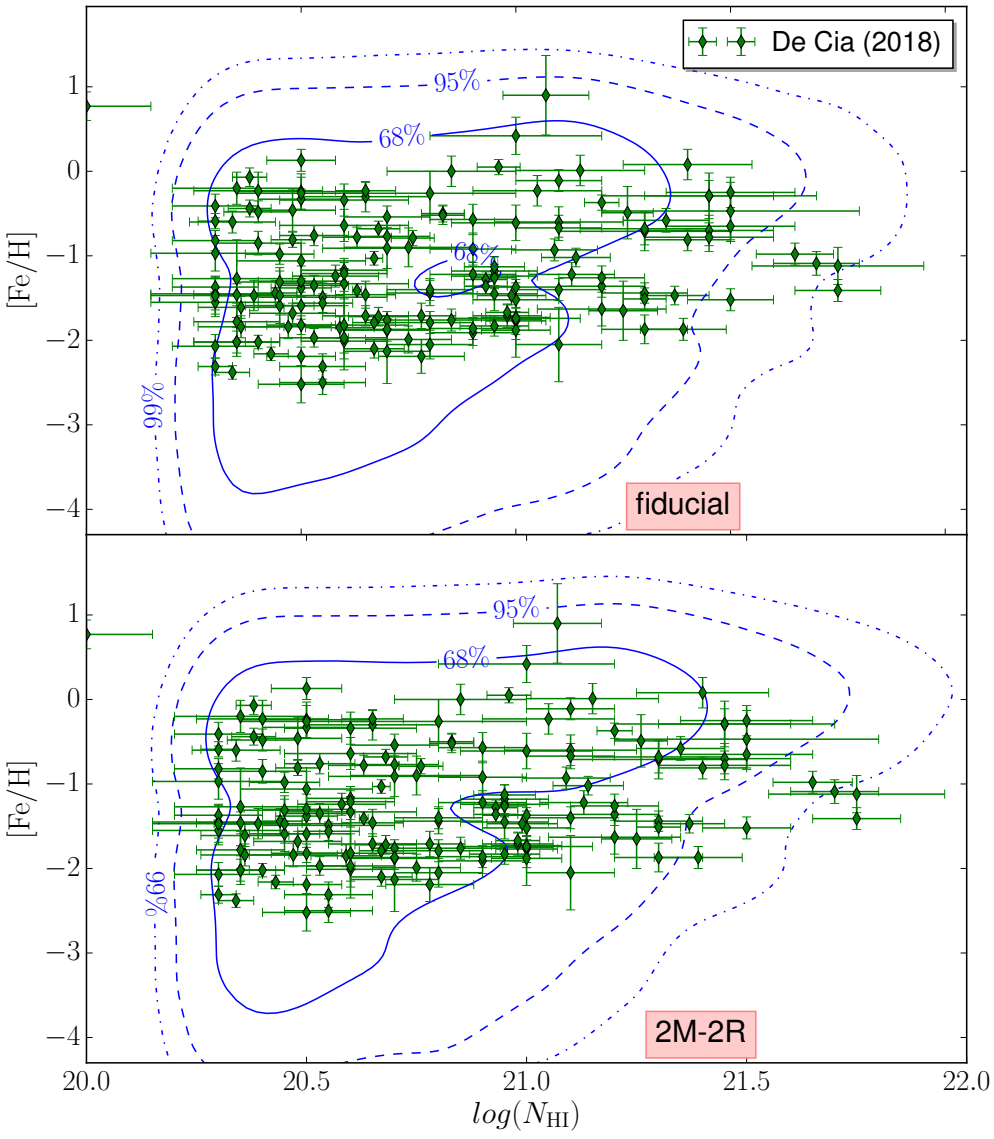

Fig. 3.10: $[\frac{\rm Fe}{\rm H}]$ as a function of $N_{\rm HI}$ in the redshift range $2 < z < 3$, with no correction for a metallicity gradient. The top and bottom panels show the metallicity distributions based on the fiducial and 2M-2R models, respectively. In both panels, we show the distributions of abundance ratios obtained by stacking the 125 realizations considered.

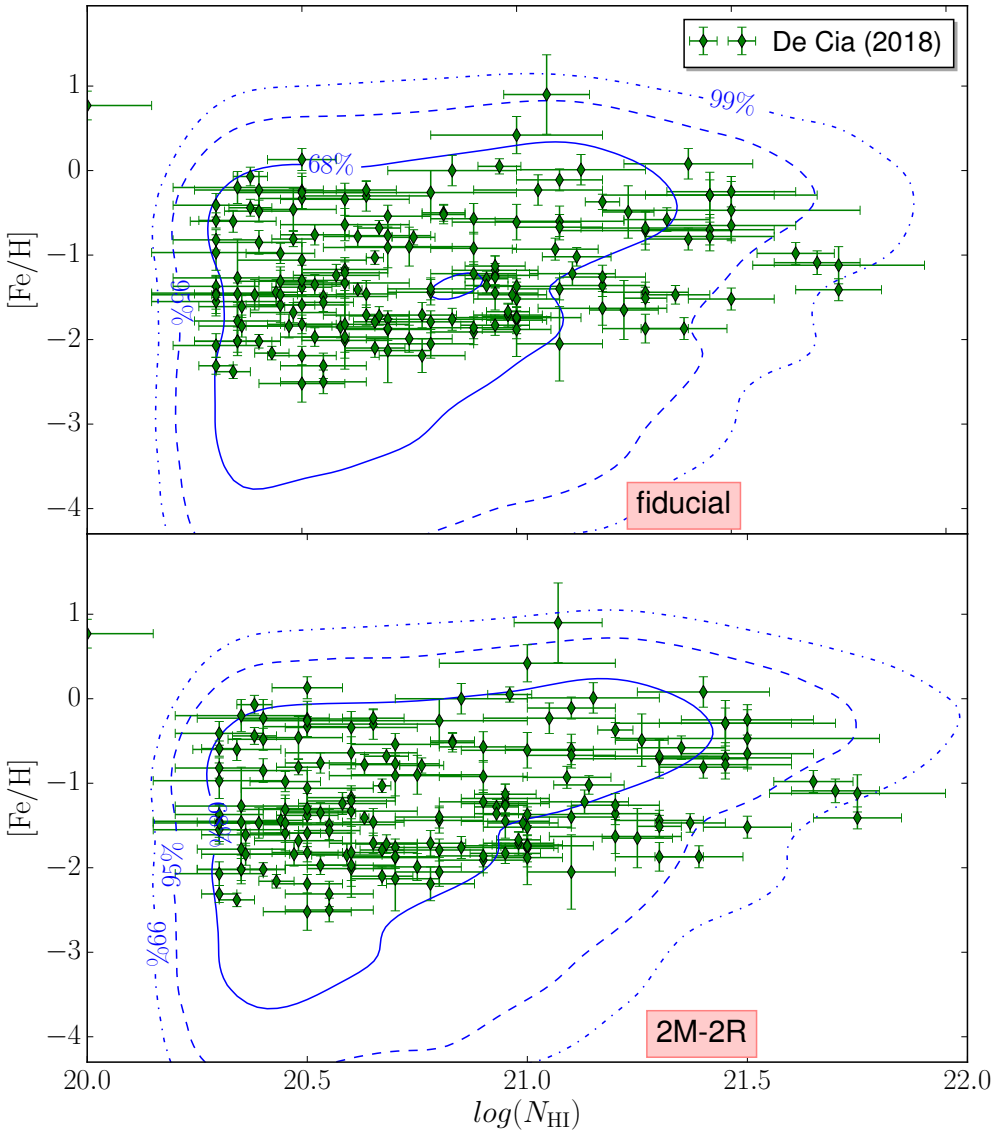

Fig. 3.11: As in Fig. 3.10, but applying a correction for the metallicity gradient, based on the fitting formula by Christensen et al. (2014b).

the mean cosmic metallicity of simulated DLAs at different redshifts. Following Rafelski et al. (2012), this can be defined, at each redshift, as:

$$< \Omega_Z >= log_{10}(\sum_i \frac{10^{[M/H]_i} N_{HI,i}}{\sum_i N_{HI,i}}) \quad (3.11)$$

where the index i runs over all DLAs in the redshift bin considered, and $[M/H]_i$ is the adopted metal abundance ratio (in our case [Fe/H]). Fig. 3.12 shows the mean cosmic metallicity as defined in Eq. 3.11 for the redshift range $0.3 < z < 4$, together with the fitting function (black dashed line) found by Rafelski et al. (2012) and the data from De Cia et al. (2018). We consider the latter sample as our reference data sample, since our metal abundances do not account for dust depletion.

When we apply a correction for the metallicity radial gradient, the mean evolution of the cosmic metallicity of our simulated DLAs is in agreement with the data by De Cia et al. (2018) within the errors, although model predictions tend to give always higher median values than the median of the data, at all redshifts considered. This is expected because, as discussed above, observations likely miss most of the DLAs at high column densities with large metallicity.

3.4.4 DLA host halo masses

The typical range of halo masses hosting DLAs is still an open question, albeit the low detection rate of DLA counter-parts in optical follow-up observations suggest that DLAs are most likely associated with faint galaxies and therefore reside in small haloes (Fynbo et al., 1999; Krogager et al., 2017). If we adopt the same argument used in abundance matching studies (e.g. Conroy et al., 2006), the results by Fynbo et al. (2008) can be translated into a typical DLA host halo mass of $M_{200} < 10^{11}\, M_{\odot}$. This is, however, in tension with more recent observational work based on DLA kinematics and clustering. The distribution of velocity widths measured from low ionization metal lines shows a prominent tail at high velocities, which suggests the existence of a population of large discs hosting DLAs (Bird et al., 2015). Moreover, the recent cross-power spectrum analysis by Font-Ribera et al. (2012), based on the BOSS survey, provides an estimate of the linear bias of the observed DLAs ($\mathrm{b_{DLA}} = 2.17 \pm 0.20$), suggesting a typical host halo mass $\sim 10^{12} M_{\odot}$. This analysis has been updated by Pérez-Ràfols et al. (2018a), who found a linear

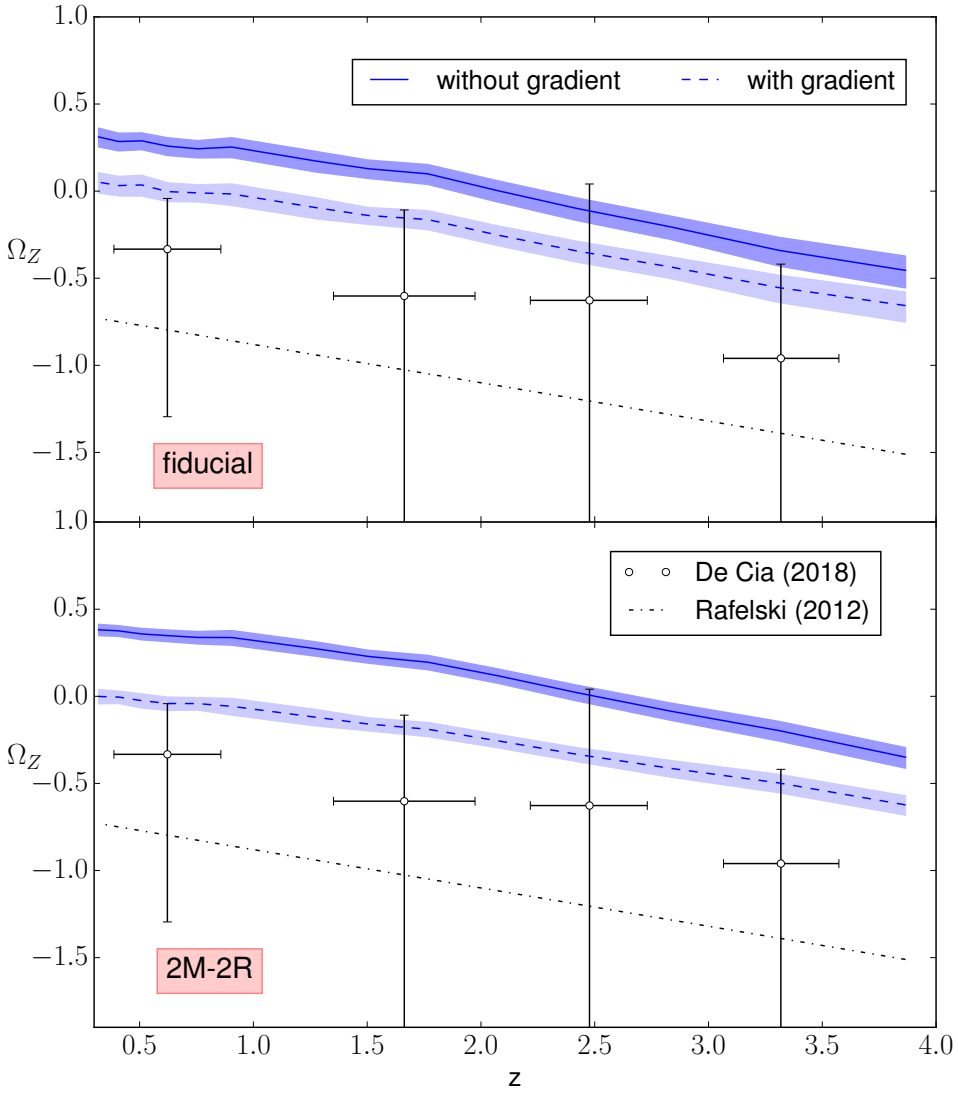

Fig. 3.12: Cosmic metallicity evolution. Solid and dashed blue lines show model predictions without and with a correction for metallicity gradient, respectively. The shaded areas highlight the relative $1-\sigma$ scatter regions. The top panel shows results for our fiducial GAEA run, while the bottom panel corresponds to the 2M-2R model.

z	M_{200} (fiducial)	M_{200} (2M-2R) $[10^{11} M_{\odot}]$
2.07	$1.50 \pm^{0.64}_{0.21}$	$2.67 \pm^{0.56}_{0.50}$
2.42	$1.32 \pm^{0.14}_{0.23}$	$2.00 \pm^{0.38}_{0.43}$
3.06	$0.71 \pm^{0.18}_{0.11}$	$0.94 \pm^{0.21}_{0.11}$

Table 3.1: Median DLA host halo masses predicted by our fiducial and 2M-2R model.

bias of $\mathrm{b_{DLA}} = 2.00 \pm 0.19$, only slightly lower than the clustering amplitude measured for Lyman Break Galaxies (LBGs, see Cooke et al., 2006), and no dependence of the bias value on redshift or column density. This bias value implies $M_{\mathrm{host}}^{\mathrm{DLA}} \geq 10^{11} M_{\odot}$, that is larger than the one typically predicted by some simulations and semi-analytic models (Pontzen et al., 2008; Barnes et al., 2014; Padmanabhan et al., 2017). The median typical DLA host halo masses, found in the redshift range $2 < z < 3$ from our model, are listed in Table 3.1 and are in agreement with observational results by Pérez-Ràfols et al. (2018a).

Recently, Pérez-Ràfols et al. (2018b) have shown that the bias of DLAs exhibits a dependence on metallicity, in line with preliminary observational results (Neeleman et al., 2013; Christensen et al., 2014b) and the expectation that more metal-rich DLAs are associated with more massive galaxies. In our model, we also see a variation of the average metal content of DLAs hosted in haloes of different masses, and this can be explained as a consequence of the relation between the gas metallicity in galaxies and the host halo mass (see Fig. 3.7).

3.4.5 Evolution of the DLA statistics with redshift and in different mass bins

Observations indicate a negligible evolution of the CDDF as a function of redshift (Noterdaeme et al., 2012), while other theoretical studies based on hydrodynamical simulations (Rahmati et al., 2013) found little evolution of the low column density end, with the slope becoming steeper at higher redshift. In contrast with observations, our model predicts a moderate evolution of the CDDF, in particular of the low column density end, that flattens at lower z. To understand the origin of this evolution it can be useful to investigate how different DLA host halo masses are distributed in different

column density bins.

Fig. 3.13 shows the predicted CDDF at three different redshifts ($z = 2.83,\ 2.42,\ 2.07$), with the dashed lines highlighting the contribution of haloes of different mass, and the bottom (top) panel showing the results of our $2M - 2R$ (fiducial) model.

Haloes in the mass bin $10^{11} M_{\odot} \leq M_{200} < 10^{12} M_{\odot}$ represent the major contribution to the CDDF at all column densities - a contribution that decreases at higher redshift, as expected in a hierarchical scenario. The second major contribution come from haloes in the mass bin $10^{10} M_{\odot} \leq M_{200} < 10^{11} M_{\odot}$ for intermediate/low DLA column densities and from haloes in the mass bin $10^{12} M_{\odot} \leq M_{200} < 10^{13} M_{\odot}$ for high DLA column densities. Haloes with $M_{200} < 10^{10} M_{\odot}$ start to contribute significantly for $log(\mathrm{N_{HI}}) < 20.7$ around $z \sim 2.8$ and moving to higher redshift their relative contribute to all column densities increases.

In the $2M - 2R$ model the contribution to the CDDF of the haloes with $M_{200} > 10^{11} M_{\odot}$ increases at all column densities while the contribution of low mass haloes ($M_{200} < 10^{10} M_{\odot}$) increases only at low column densities (for $N_{\mathrm{HI}} < 21$) and that of the intermediate mass ($10^{10} M_{\odot} < M_{200} < 10^{11} M_{\odot}$) haloes increases for intermediate column densities (up to $N_{\mathrm{HI}} = 1.4$), with respect to the fiducial model. This could be explained remembering that the HI surface density scales linearly with the mass and as the inverse of the square radius of the galactic disk. Moreover, since the cross-section increases quadratically with the galaxy scale radius, a larger number of disks are intersected by our l.o.s. in the 2M-2R model, in particular at larger halo masses. Therefore, the predicted CDDF from this model is in better agreement with observational measurements.

In Fig. 3.14 we show, for the redshift range $0 < z < 4$ the contributions to the predicted $\Omega^{\mathrm{HI}}_{\mathrm{DLA}}$ of haloes in different mass bins, with logarithmic bin size $\Delta \log(M_{200}/M_{\odot}) = 1$. The haloes which contribute more at all redshifts are those in the mass bin $10^{11} M_{\odot} \leq M_{200} < 10^{12} M_{\odot}$. The second largest contribution is provided by the mass bin $10^{12} M_{\odot} \leq M_{200} < 10^{13} M_{\odot}$ up to $z = 2.5$ and by the mass bin $10^{10} M_{\odot} \leq M_{200} < 10^{11} M_{\odot}$ for $z > 2.5$. It is worth noting that the contributions of the two lowest mass bins are very similar and both represent less than 10% of the total $\Omega^{\mathrm{HI}}_{\mathrm{DLA}}$ in the entire redshift range considered.

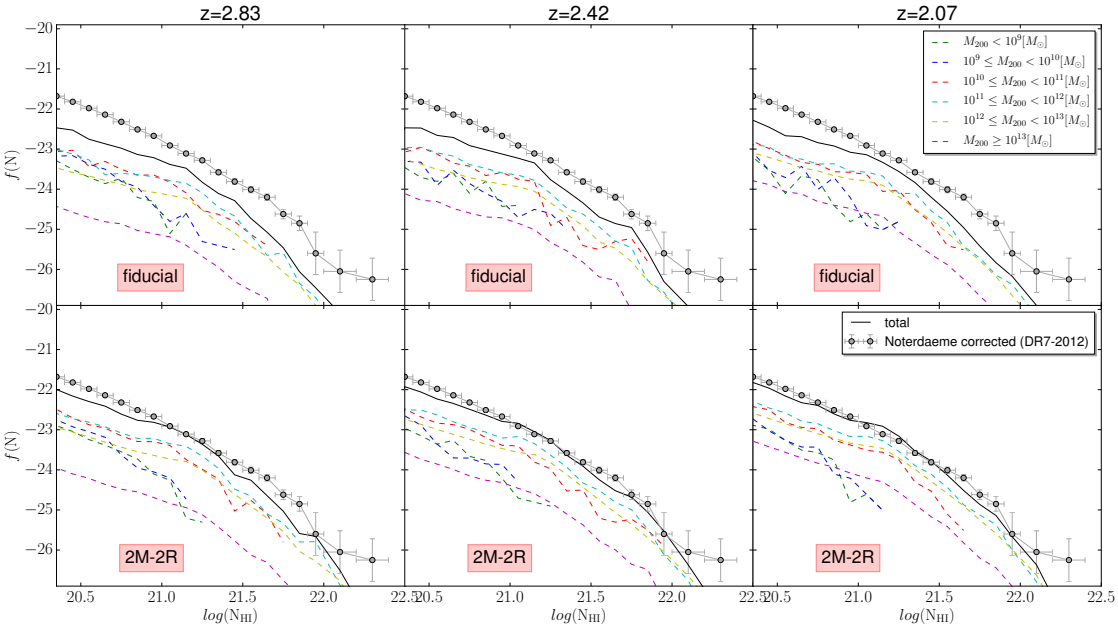

Fig. 3.13: Evolution with redshift of the predicted CDDF and its dependence on the DLA host halo masses. The top panels show results from our fiducial model, while the bottom panels show the corresponding results from the 2M-2R model. The black solid lines show the total CDDF, while the dashed colored lines show the average contribution to the CDDF of dark matter haloes in different virial mass bins, as indicated in the legend.

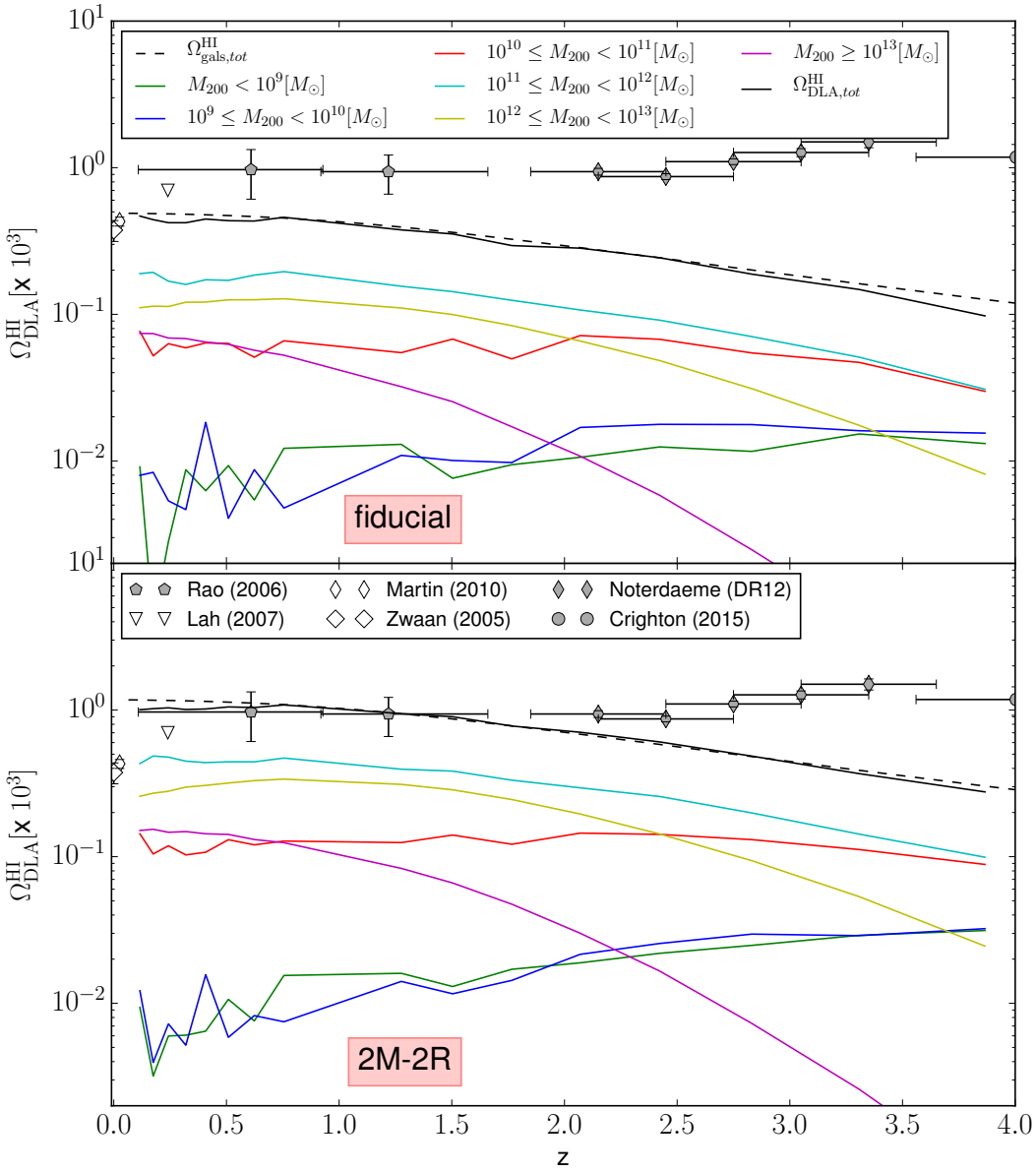

Fig. 3.14: Evolution with redshift of the predicted comoving HI density in DLAs ($\Omega_{\rm HI}^{\rm DLA}$) and its dependence on the DLA host halo masses. The top panel shows results from our fiducial model, while the bottom panel shows the corresponding results from the 2M-2R model. The black solid lines show the total $\Omega_{\rm HI}^{\rm DLA}$, while the dashed colored lines show the average contribution to $\Omega_{\rm HI}^{\rm DLA}$ of dark matter haloes in different virial mass bins, as indicated in the legend.

3.4.6 Distribution of impact parameters versus $\mathrm{N_{HI}}$

For our model galaxies, we define the impact parameter, b, as the distance between the LOS and the center of mass of the galaxy hosting the DLA. In observations, b measures the projected distance between the (luminosity) center of the galaxy and the quasar sight-line piercing the cold gas.

In Fig. 3.15, we show the distribution of impact parameters as a function of the hydrogen column density ($\mathrm{N_{HI}}$), obtained considering DLAs originating from the MSI and MSII haloes, in the redshift range $2 < z < 3$ and with $\mathrm{[Fe/H]} > -2$. Model predictions are compared with observational measurements by Krogager et al. (2017), that cover the same redshift and metallicity range. The data point come partly from the literature and partly from an X-shooter follow-up campaign. The latter is the first sample of DLA counterparts at high redshift associated with a relatively high detection rate ($\sim 64\%$), likely due to the adopted DLA pre-selection: $\mathrm{EW_{SiII}} > 1$Å($\mathrm{EW_{SiII}}$: rest-frame equivalent width of the $\mathrm{Si_{II}}$ line, with $\lambda = 1526$Å) implying large metallicities.

The top panel of Fig. 3.15 shows results from the run of GAEA described in 3.2.2, our fiducial model, while the bottom panel shows results obtained multiplying by a factor 2 both the scale radius and the cold gas mass of all model galaxies. In the following, we will refer to this as the $2M - 2R$ model.

The largest 99 per cent contour level of the simulated distribution, in both the fiducial and the $2M - 2R$ model, encloses all the observed data. In the $2M - 2R$ model, all data points fall inside the 95 per cent contour level of the simulated distribution (for MSI haloes), and there is a more clear anti-correlation between impact parameters and column density. For the fiducial model, we find $< b >^{sim}_{DLA} = 8.23$ (3.00) for MSI (MSII). The corresponding values for the 2M-2R model are $< b >^{sim}_{DLA} = 14.63$ (5.14). The different mean value of the impact parameters and the different contour levels between DLA originated from the MSI and MSII haloes reflect the dependence of the galactic disk size on the virial radius of the halo where the galaxy resides.

Averaging and weighting over the relative contribution of MSI and MSII, we obtain $< b >^{sim}_{DLA} = 5.53$ and $< b >^{sim}_{DLA} = 10.03$ for the fiducial and the $2M - 2R$ model, respectively. Both estimates are in agreement with the one found by Krogager et al. (2017), $< b >_{DLA} = 8.32$, with a slight preference

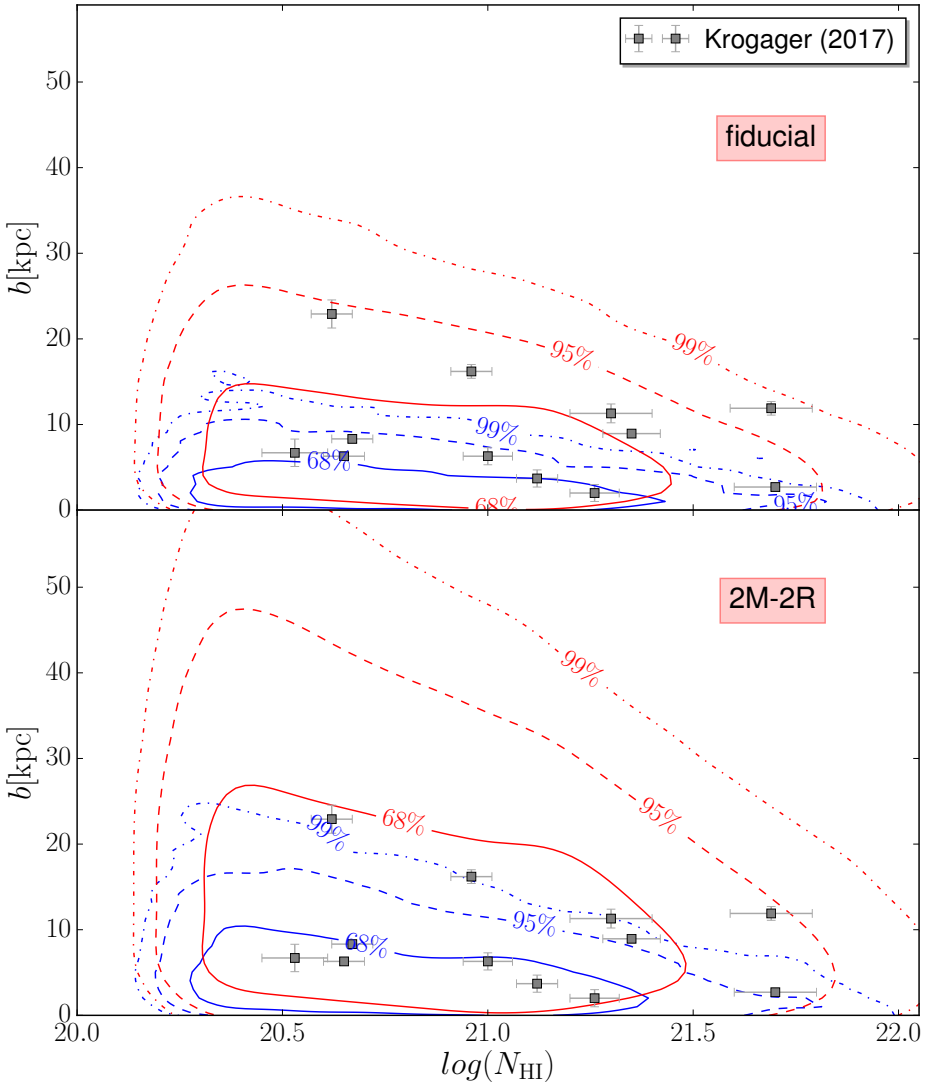

Fig. 3.15: Impact parameter (b) as a function of the hydrogen column density N_{HI}, for simulated DLAs in the redshift range $2 < z < 3$, compared to the observations by Krogager et al. (2017, grey symbols). The top panel shows the contour distributions based on the fiducial model, while the bottom panel shows results obtained multiplying by a factor 2 both the scale radius and the cold gas mass of all model galaxies ($2M - 2R$ model, hereafter). Red and blue lines refer to galaxies in the MSI and MSII, respectively, and show different contour levels of the distribution as indicated in the legend. We apply a cut to the metallicity of model DLAs equal to $[Fe/H] > -2.0$, for consistency with the observational measurements considered.

for the $2M - 2R$ model.

Extending the sample of observed DLAs towards lower redshift, Rhodin et al. (2018) reports $< b >_{DLA} = 11.1$ kpc. Older work by Rao et al. (2011b) based on low-redshift DLAs counterparts found $< b >_{DLA} = 17.4$ kpc, considering a larger metallicity cut ($[\mathrm{Fe/H}] > -1$). The different observational estimates depend on the adopted DLA pre-selections, on the techniques used to search for DLAs counterparts, and in part also on the expected redshift evolution of galaxy sizes that implies an evolution of the observed range of impact parameters.

It is important to bear in mind that all observations of DLA counterparts are biased against smaller impact parameters, for which it is difficult to detect the DLA counterparts (as discussed in Krogager et al., 2017). The technique adopted by Krogager et al. (2017) likely misses also some counterparts at large impact parameters due to the partial coverage of the FoV by the three long-slits while DLA systems with very high metallicities are not completely detected because of the dust bias (Khare et al., 2012), which affects the colour selection of QSOs. Therefore, the comparison shown in Fig. 3.15 should be considered more as qualitative than rigorous.

3.5 Discussion of the results

In this work, we have analysed the properties of Damped Lyman$-\alpha$ systems (DLAs) by taking advantage of a semi-analytic model of galaxy formation and evolution (GAEA, presented in Hirschmann et al., 2016; Xie et al., 2017) coupled to two large cosmological N-body simulations: the Millennium (MSI) and Millennium II (MSII). In order to estimate the possible contribution from haloes that are below the resolution of our simulations, we have used a simple HOD approach by placing, at random positions within the simulated box, a number of haloes with mass distribution consistent with analytic formulations tuned on N-body simulations (Tinker et al., 2008). Our model assumes that all atomic hydrogen is associated with the gaseous disk of galaxies, i.e. there is no contribution from filamentary regions or extraplanar gas. Our simulated DLAs catalogues are then built by throwing a large number (100,000) of random lines of sight along the z-direction of 125 simulated boxes, obtained combining the simulations available and complemented with HOD extrapolation (as described in Section 3.3).

Our fiducial model predicts a column density distribution function with the correct shape but offset low with respect to observational measurements by Noterdaeme et al. (2012). This affects the predicted values of the cosmic hydrogen density in DLAs ($\Omega^{\rm HI}_{\rm DLA}$), that is a factor ~ 2.5 lower than observational estimates at $0 < z \leq 2$, and even more at higher redshift. Up to $z < 3$ the disagreement with data can be overcome by increasing the radius of the gaseous disk and the gas mass by a factor ~ 2 (our $2M - 2R$ model). As for the DLA metallicity distribution in the redshift range $2 < z < 3$, our model predicts an excess of low metallicity DLA systems, while the average cosmic DLA metallicity ($\Omega_{\rm Z}$), weighted over $\rm N_{HI}$, follows the same redshift evolution as observational measurements but it is slightly higher than observed values. The predicted $\Omega_{\rm Z}$ becomes compatible with observations, within the uncertainties, once we account for a modest radial metallicity gradient.

Below, we discuss our results in relation with independent recent studies, and point out possible developments/improvements of the adopted physical model that can bring model results in better agreement with observational measurements.

3.5.1 Comparison with the literature

In the last twenty years, a number of theoretical studies, either using a semi-analytic approach (Lagos et al., 2011, 2014; Berry et al., 2014; Kim et al., 2015) or hydro-dynamical simulations (Nagamine et al., 2007; Pontzen et al., 2008; Tescari et al., 2009; Altay et al., 2011; Cen, 2012), have focused on the evolution of the atomic hydrogen content of the Universe.

In the framework of this paper, it is particularly interesting to discuss our results in relation to the analysis by Berry et al. (2014, 2016), also based on semi-analytic models and focused on the predicted properties of DLAs. In their work, Berry et al. (2014) use variations (see their Table 1) of the semi-analytic model published in Somerville et al. (2008, see also Popping et al. 2014; Somerville et al. 2015), including different prescriptions for the partition of cold gas in atomic and molecular hydrogen, and alternative assumptions for the sizes of gaseous disks. Our model and the one used by Berry et al. (2014) differ significantly for the numerical implementation and for the prescriptions adopted for modelling various physical processes. Popping et al. (2014) and Xie et al. (2017) show that both models are able

to reproduce the evolution of disc sizes (both stellar and gaseous) up to $z \sim 2$, for galaxies more massive than $10^9 M_{\odot}$. Fig. 2 of Berry et al. (2014) shows that none of the model variants they considered reproduces well the local HI mass function, while our fiducial model is tuned to reproduce this observational constraint.

Both our fiducial run and the reference disc model used in Berry et al. (2014) under-predict the column density distribution function of DLAs. Berry et al. (2014) find a better agreement by increasing the cold gas specific angular momentum with respect to what assumed in their reference model. This leads to larger gaseous disks, but also to a significantly worse agreement with the HI galaxy mass function in the local Universe (see their Fig.2). This is consistent with our findings that a model where we arbitrarily multiply by a factor two both the scale radii and HI masses of model galaxies better reproduces the observed column density distribution function. Ours is an 'ad-hoc' solution, and it remains to be demonstrated that plausible modifications of the modelled physical processes can lead to such solution without (significantly) affecting the agreement shown between model predictions and observational data in the local Universe. We will come back to this issue in the next section, in the framework of possible developments of the GAEA model.

It should be noted that also our 2M-2R model, that reproduces the observed column density distribution of DLAs for $z < 3$, predicts a decline of $\Omega^{\rm HI}_{\rm DLA}$ at higher redshift. This is in disagreement with observational measurements and consistent with what found by Berry et al. (2014, 2016). This decline is driven by an under-estimation of the column density distribution function for $log(N_{\rm HI}) < 21$. The behaviour is not shared by hydro-dynamical simulations that typically not underestimate the CDDF for $log(N_{\rm HI}) < 21$ and find no evolution or even a moderate increase of $\Omega^{\rm HI}_{\rm DLA}$ (Cen, 2012; van de Voort et al., 2011; Altay et al., 2011), in better agreement with observational measurements.

The different behavior at high redshift, predicted by semi-analytic models and hydro-dynamical simulations, could be at least in part explained by an increasing contribution to the DLAs cross-section of filamentary structures and outflows/inflows at higher redshift (van de Voort et al., 2011; Fumagalli et al., 2011; Cen, 2012). In addition, simulations predict that at $z \sim 3$ the halos that contribute most to the CDDF for $log(N_{\rm HI}) < 21$ are the ones

in the mass range $10^9 < M_{200} < 10^{10} M_\odot$ (Tescari et al., 2009; Rahmati et al., 2013) while in our model the major contribution comes from halos in the mass range $10^{10} < M_{200} < 10^{12} M_\odot$. Since more massive halos are less numerous at higher redshift, the difference in the typical DLA host halo mass at $log(N_{\rm HI}) < 21$ could partially explain the decline of the $\Omega^{\rm HI}_{\rm DLA}$ in our model. Another concern is related to the possible contribution of haloes that are below the resolution of our simulations. In order to understand to what extent low-mass haloes contribute to the HI comoving density, we have estimated the contribution of haloes with mass $10^8 {\rm M}_\odot < {\rm M}_{200} < 10^{9.2} {\rm M}_\odot$ resorting to a simple HOD model (see Sec.3.2 for details). Our results indicate that these low-mass haloes represent a negligible contribution to the column density distribution in the redshift range of interest. The average covering fraction of HI in different halos is influenced also by the interplay between the UV background and the gas density in the galactic disks. At the column density typical of DLA systems, the gas is self-shielded by the ionizing photons of the UV background, then mostly neutral. Our semi-analytical model does not include a specific treatment for the self-shielding but this effect is taken into account implicitly through the adoption of the BR prescription for the cold gas partitioning (Blitz & Rosolowsky, 2006). Albeit the Blitz & Rosolowsky (2006) prescription is based on observations of local galaxies, we are confident in applying this prescription to all redshifts, since we have demonstrated that it provides very similar results to alternative parametrizations based e.g. on hydro-dynamical simulations that account explicitly for self-shielding (Xie et al., 2017).

As discussed in the previous sections, both our fiducial and $2M - 2R$ models predict an excess of low-metallicity DLAs that are not present in observational samples, also in the dust-corrected DLA abundance ratio catalog by De Cia et al. (2018). Results based on the model by Somerville et al. (2015) appear in better agreement with the observed metallicity distribution of DLAs (see Fig.10 in Berry et al., 2014). This difference is likely due to the different treatment adopted for the metal enrichment. In particular, the Somerville model assumes an instantaneous recycling approximation and sets a metallicity floor for the hot gas in low mass haloes (the haloes with $M_{vir} \leq 10^9 M_\odot$ are set to have a hot gas metallicity equal to $10^{-3} Z_{solar}$). Our model instead does not assume pre-enrichment of gas in low-mass haloes and includes a detailed chemical enrichment scheme that accounts

for the non instantaneous recycling of gas and metals (De Lucia et al., 2014). In addition, as discussed above, we assume that the 95% of newly synthesized metals is directly injected into the hot gas phase in low mass haloes (Hirschmann et al., 2016), which contributes to delay the chemical enrichment of low-mass systems. Fig. 16 of Somerville et al. (2015) shows that their model predicts almost no evolution with redshift of the mass-metallicity relation while our model predicts an increasing normalization at lower redshift. The different redshift evolution together with the different slope (less steep for low-mass galaxies in the case of the Somerville model) could lead to gas metallicities, for galaxies in the mass range $10^7 < M_\star < 10^8$, that are larger in the Somerville et al. model than in ours. This could also contribute to the different predictions obtained for the metallicity distribution of DLAs, in particular at low metallicities.

For the typical DLA host halo mass, predictions from our model (both for the fiducial and the $2M - 2R$ run) are similar to those by the Somerville model. For a mean redshift $z = 2.3$ (taking all DLAs with $1.97 \leq z \leq 2.6$) we find a median DLA host halo mass equal to $M_{host}^{DLA} = 1.55 \times 10^{11} M_\odot$ for the fiducial model and $M_{host}^{DLA} = 2.28 \times 10^{11} M_\odot$ for the $2M - 2R$ model, in agreement with observational estimates by Pérez-Ràfols et al. (2018a).

3.6 Conclusions

Our results suggest possible avenues to improve the agreement between the predictions of the GAEA model and observational data of DLAs: (i) increasing the HI content of model galaxies; (ii) increasing the sizes of gaseous disks; and (iii) modifying the treatment for the metal enrichment of low-mass haloes. In this section, we discuss plausible implementations that can bring the model in this direction. In future work, we intend to explore these suggestions in more detail.

The HI content of model galaxies depends on the assumed prescription for cold gas partitioning. In our model, the molecular to atomic hydrogen ratio is slightly larger than what observed in the local Universe by xGASS and xCOLDGASS (Catinella et al., 2018; Saintonge et al., 2017). This is shown and discussed in a forthcoming paper (Xie et al., 2020a). Naively, one could think that a lower molecular fraction can be obtained by simply increasing the star formation efficiency: stars are formed from molecular gas and larger

star formation rates should lead to consume more molecular hydrogen. The situation is, however, complicated by the strong self-regulation between star formation and stellar-feedback that makes model results not very sensitive to the star formation law adopted (Xie et al., 2017, and references therein). In addition, simple modifications of model parameters would generally require a retuning of the model to restore the agreement with the main observables used as constraints (in our case the HI mass function).

Another possible reason for the too low HI masses of galaxies in intermediate mass haloes is the prescription adopted for reionization. Our model assumes an 'early' reionization (with starting redshift $z_0 = 15$ and completed by $z_r \sim 11$) that is inconsistent with recent Planck results (Planck Collaboration et al., 2018). The reionization feedback is implemented through a 'filtering mass' whose evolution is described by the analytic fitting function introduced in Kravtsov et al. (2004) (based on the simulation results by Gnedin, 2000). Adopting a time-line for reionization more in agreement with recent results, we expect a filtering mass lower by an order of magnitude with respect to the one assumed in our model for $z > 5$. Besides, Okamoto et al. (2008b) showed that the parametrization of the filtering mass presented in Gnedin (2000), based on low-resolution simulations, might over-estimate by up to one order of magnitude (at $z = 0$) the characteristic mass where photo-ionization feedback becomes effective in reducing the baryon fraction (see their Fig. 6), independently of the assumed reionization history.

The size of the HI galactic disks in our model is determined by the evolution of the specific angular momentum of the cold gas. Xie et al. (2020a) have significantly updated the treatment of the angular momentum, leading to both larger gaseous disks and larger HI masses, in the direction of the 2M-2R model considered in previous sections. It is worth noting that also the SFR sizes predicted by our fiducial model tend to be smaller than observational estimates (Xie et al., 2017), and also this disagreement is relieved with the larger disks obtained with the updated angular momentum scheme (as shown in, Xie et al., 2020a).

Finally, the excess of low-metallicity DLAs, with respect to observational measurements, can possibly be solved by modifying the fraction of metals that are injected directly into the hot gas component in low-mass haloes. This will likely affect also the cooling times (the cooling function is very sensitive to the metallicity of the hot gas), leading to a lower gas accretion rate onto

galaxies, but could be compensated by the above-described modifications concerning the reionization scenario.

Chapter 4

Testing prescriptions for a better HI modelling in the GAEA model

Adopting as a starting point the results of the work presented in Chapter 3, I dedicated the last year of my PhD to improve the modelling of HI in the framework of the GAEA model, using in particular DLA observations as a test set. I worked on a new version of the model (Xie et al., 2020b), that includes a detailed treatment for the ram pressure stripping of cold gas and implements updated modelling to trace the specific angular momentum of the gaseous disks in galaxies. I will refer to this model version as the *RPS model*.

In the framework of the RPS model three types of improvements have been tested :

1. the assumed reionization scenario was changed from an early to a late reionization, in agreement with recent constraints derived from the Planck mission (Planck Collaboration et al., 2018) and other high-z reionization probes (Greig & Mesinger, 2017);
2. different implementations of the reionization feedback or the UVB feedback (in particular the ones described in Macciò et al., 2010; Kim et al., 2015);
3. different schemes for the partitioning of the ISM cold gas phase into its atomic and molecular components are described in detail in section 4.4.

4.1 How are treated Reionization and Reionization feedback in the default GAEA model?

Before Reionization started, the galaxy formation in a halo simply needed that the virialized halo gas had sufficient time to cool and condense. During the Reionization epoch the UV photons, which reionized the IGM, caused also a photo-heating of the baryons in the IGM to $\sim 10^4$ K, leading to the suppression of gas accretion onto these objects (Thoul $\&$ Weinberg 1996(Gnedin, 2000; Hoeft et al., 2006; Okamoto et al., 2008b; Noh & McQuinn, 2014; Katz et al., 2019).

The photo-ionization feedback, due to the same UV photons which reionized the Universe, is included in the GAEA semi-analytical model through the introduction of a filtering halo mass, below which the accreting baryon fraction is reduced. Our implementation for the filtering mass is based on the analytical formula for the filtering mass evolution with redshift by Kravtsov et al. (2004), derived from the simulation results by Gnedin (2000).

The Gnedin's work, published in 2000 and based on two SPH simulations at high resolution (with very small box sizes, respectively of 4 and 2 Mpc), similarly to what is done in the other first theoretical works that treated the suppression of galaxy formation due to UV feedback (Shapiro et al. 1994, Gnedin $\&$ Hui 1998), relies on a comparison between the effective mass under which an halo is subject to reionization feedback and the characteristic scales of linear and non-linear theory of structure formation [1]. In Fig. 3 of his article (reported here for clarity) Gnedin compared the redshift evolution of the average halo mass of the haloes with a baryon fraction $f_b =\sim 0.5$ (half of the universal baryon fraction) with the linear Jeans mass and the Jeans mass at the virial overdensity, showing that neither the linear or the non-linear Jeans mass reproduce the evolution of the characteristic mass resulting from the Reionization feedback. Instead, he found that the filtering mass (M_F) can be a good approximation for this characteristic mass (M_c), fitting the redshift evoultion of the gas mass associated to different halo masses with the following formula:

$$\bar{M_g}(M_t, t) = \frac{f_b M_t}{[1 + (2^{\alpha/3} - 1)(M_c(t)/M_t)^{\alpha}]^{3/\alpha}} \tag{4.1}$$

[1] e.g Jeans mass in linear and non linear theory, or the filtering mass

where $\alpha = 1$ gives the best values for the χ^2 test at all redshifts.

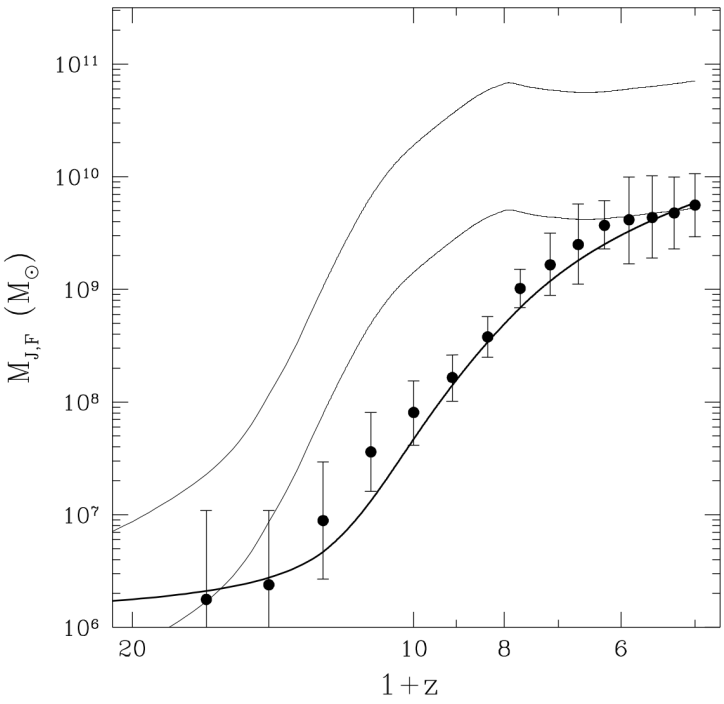

Fig. 4.1: Fig. 3 from Gnedin (2000). Redshift evolution of different typical mass scales for halo growth. The two thin lines show the linear Jeans mass (top) and the Jeans mass at the virial overdensity of 180 (bottom). The bold line shows the filtering mass ($M_{J,F}$), and the filled circles with error bars show the median values for the characteristic mass at which $M_b = 0.5 f_b M_h$, as measured from the simulations.

Therefore, Gnedin (2000) defined $M_{\rm F}$ as the mass of the halo which would loose half of the baryons, compared to the universal baryon fraction. This *filtering mass* relates to the Jeans mass of the intergalactic gas integrated over the cosmic history as described in the equation below (eq. 6 in Gnedin, 2000):

$$\begin{aligned} M_{\rm F}(a) &= M_{\rm J0}\, f(a)^{3/2}, \\ f(a) &= \frac{3}{a}\int_0^a x\, T_4(x) \left[1 - \left(\frac{x}{a}\right)^{1/2}\right] dx \end{aligned} \tag{4.2}$$

where $M_{\rm J0} = 2.5 \times 10^{11} h^{-1} \Omega_0^{-1/2} \mu^{-3/2}\, M_\odot$, $\mu \approx 0.59$ is the mean molecular weight of the fully ionized gas, and the integration extends over the expansion factor (a). The temperature of the cosmic gas T_4 is expressed in units of 10^4 K for convenience.

Kravtsov et al. (2004) proposed an analytical fit to the results of Gnedin (2000), assuming a simple dependence of the temperature on the expansion factor (a), dividing the Reionization epoch in three distinct stages:

$$T_4(a) = \begin{cases} (a/a_o)^{\alpha}, & \text{for } a \leq a_o \\ 1, & \text{for } a_o \leq a \leq a_r \\ (a/a_r)^{-1}, & \text{for } a \geq a_r \end{cases}$$

These stages correspond to the epoch preceding the formation of the first HII regions, for $z > z_o$, the epoch of the overlap of multiple HII regions, for $z_r < z < z_o$, and the epoch of complete reionization, $z < z_r$. In the first stage, before redshift $z_o \equiv 1/a_o - 1 \approx 8$, the temperature is rising as the newly-formed stars ionize their neighboring regions. The parameter α controls the rate of growth of the extragalactic UV flux; Kravtsov et al. (2004) found $\alpha = 6$ to be the best fit to their simulation results. During the overlap stage, between redshifts z_o and $z_r \equiv 1/a_r - 1 \approx 7$, the temperature is kept constant at roughly 10^4 K as the cosmic HII regions overlap. After the universe is fully ionized, at redshifts below z_r, the temperature falls adiabatically with the cosmic expansion.
With these analytical expressions for $T_4(a)$, we can integrate equation (4.2) analytically, getting the following expression for $f(a)$ in the three epochs:

$$\begin{aligned} f(a) &= \frac{3a}{(2+\alpha)(5+2\alpha)} \left(\frac{a}{a_o}\right)^{\alpha}, \quad a \leq a_o \qquad (4.3) \\ f(a) &= \frac{3}{a}\left\{ a_o^2 \left[\frac{1}{2+\alpha} - \frac{2(a/a_o)^{-1/2}}{5+2\alpha}\right] \right. \\ &\quad \left. + \frac{a^2}{10} - \frac{a_o^2}{10}\left[5 - 4\,(a/a_o)^{-1/2}\right]\right\}, \quad a_o \leq a \leq a_r \\ f(a) &= \frac{3}{a}\left\{ a_o^2 \left[\frac{1}{2+\alpha} - \frac{2(a/a_o)^{-1/2}}{5+2\alpha}\right] \right. \\ &\quad + \frac{a_r^2}{10}\left[5 - 4\,(a/a_r)^{-1/2}\right] - \frac{a_o^2}{10}\left[5 - 4\,(a/a_o)^{-1/2}\right] \\ &\quad \left. + \frac{a a_r}{3} - \frac{a_r^2}{3}\left[3 - 2(a/a_r)^{-1/2}\right]\right\}, \quad a \geq a_r. \end{aligned}$$

This analytical fit has been thought to provide an accurate modeling of the photo-heating effect of UV photons released during Reionization until

the results by Okamoto et al. (2008b) and can be easily implemented in the codes at the basis of different models. For these two reasons it has been adopted by most of the semi-analytical models dedicated to the study of galaxy formation. The analytical expression for $T_4(a)$, which depends only on two parameters $a_0(z_o)$ and a_r, allows a simple recalculation of the filtering mass for different values of the Reionization redshift. In addition, to account for the inefficiency of atomic gas cooling at $T < 10^4$ K (Gnedin & Kravtsov, 2010), in the GAEA model instead of $M_{\rm F}$ we use the maximum of $M_{\rm F}(z)$ and M_{char}, the halo mass corresponding to $T_{\rm vir} = 10^4$ K. For $T_{\rm vir} < 10^4$ K the baryons can collapse onto haloes only if they are cooled by H_2 molecules, via molecular cooling (e.g. Oh & Haiman 2002), but this process is negligible at high redshift, and it is not included in the GAEA model.

4.2 Modifications of the reionization history

The combined analysis of the CMB angular power spectra, based on the E-mode polarization measured by the Planck-high frequency instrument and CMB temperature data, of the evolution of the intergalactic Lyman-alpha opacity, based on the study of quasar spectra, and of the number density evolution of the Lyman-Alpha Emitting galaxies (LAEs), has provided constraints on the reionization history (Planck Collaboration et al., 2018; Greig & Mesinger, 2017; Greig et al., 2021). The most recent measurements imply that the reionization of the Universe was nearly completed at $z \sim 6$. The detection of complete Gunn-Peterson (GP) absorption troughs in the spectra of quasars at $z > 6.5$ suggests that the neutral fraction of the gas in the IGM increases rapidly with increasing redshift, while the rapid decline in the number density of LAEs from $z = 8$ to $z = 6$ suggests a late and fast reionization scenario.

Based on these observational constraints, I updated the reionization scenario assumed in GAEA. In the X17 version of the GAEA model it was assumed that the reionization started at $z_0 \simeq 15$ and was completed at $z_r = 11$, while now in the RPS-GAEA model I set $z_0 = 10$ and $z_r = 6$.

In the prescription for the filtering mass adopted in the GAEA model and described above, the evolution of the filtering mass, which regulates how much is suppressed, in first place, the baryonic collapse and then the galaxy formation in low-mass haloes, depends on the assumed redshift range of

reionization. Therefore, I expect to see some modifications of the predicted galaxy properties in models with different reionization histories.

In Fig. 4.2 I compare the redshift evolution of the filtering mass within an early reionization scenario (in green), like the one previously assumed in GAEA, and within a late reionization scenario (in blue).

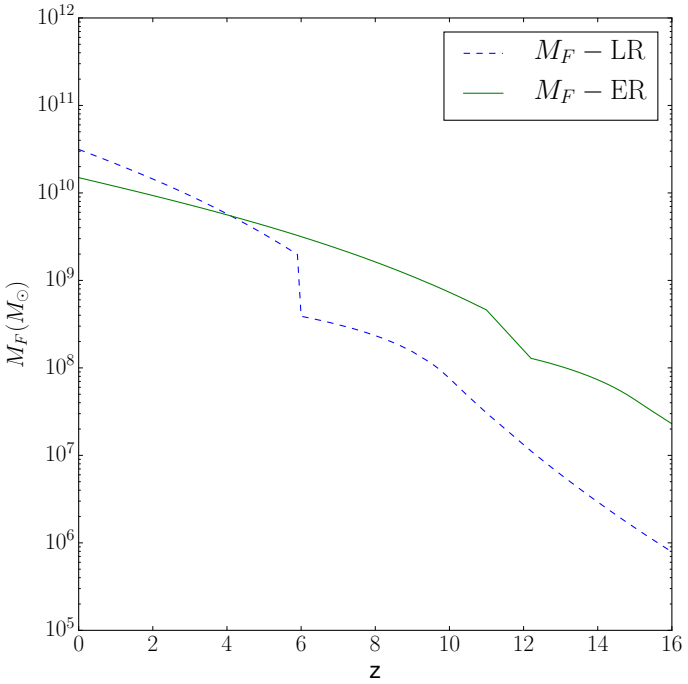

Fig. 4.2: Redshift evolution of the filtering/critical mass with different prescriptions. The labels LR and ER in the legend refer to the late and early reionization scenario respectively.

We can observe that the M_F predicted in the late reionization scenario is lower of the one predicted for the ER case down to $z \sim 5$ and become higher than the other for lower z.

The first galactic properties investigated are the cold gas/HI mass-HM relation and the SMHM relation, for central and satellite galaxies, at different redshifts. In Fig. 4.3 and Fig. 4.4, it is shown the median of the HIHM relation and SMHM relation, respectively, for central galaxies, at 4 different redshifts.

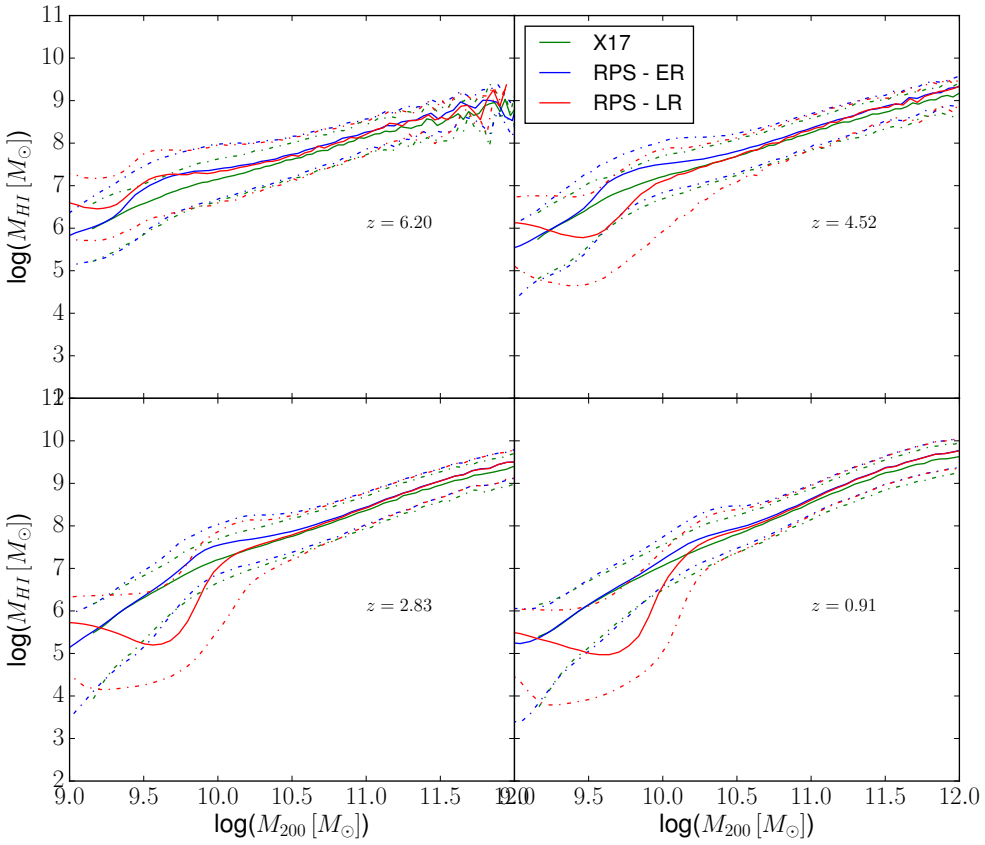

Fig. 4.3: Average HI mass for central galaxies in bins of halo mass, at $z = (6.20, 4.52, 2.83, 0.91)$. As described in the legend, in green we see the predictions by the X17 model in the early reionization scenario, and in blue/red the prediction by the RPS model in the early/late reionization scenario, respectively.

We can observe that the average HI masses of central galaxies in massive halos ($M_{200} \geq 10^{11} M_\odot$) have similar values in the X17 model (coupled to a ER scenario), in green, and in the RPS model coupled to both the two reionization scenarios, LR(red) and ER (blue), at all redshifts. At intermediate halo masses for the RPS-ER model we observe a bump, more prominent at

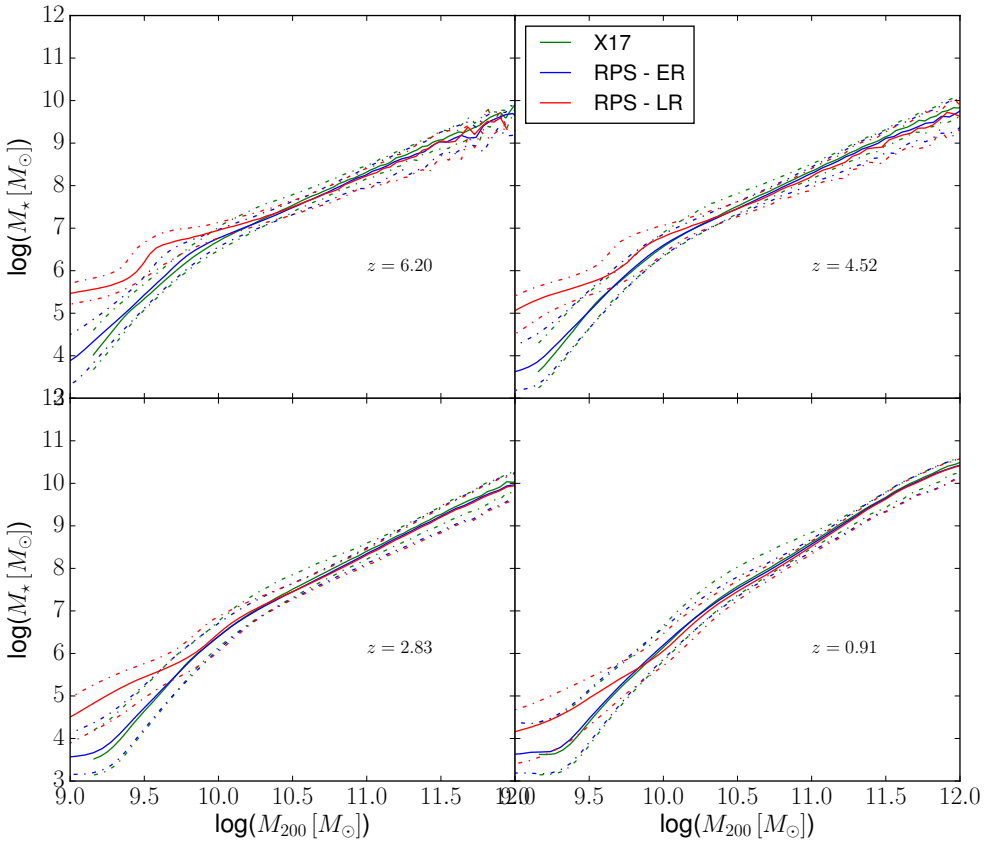

Fig. 4.4: Average stellar mass for central galaxies in bins of halo mass, at $z = (6.20, 4.52, 2.83, 0.91)$. As described in the legend in green we see the predictions by the X17 model, and in blue/red the prediction by the RPS model in the early/late reionization scenario.

high redshift, in the average predicted HI mass for central galaxies, peaking around $M_{200} = 10^{9.7} M_\odot$ at $z \sim 6$ and around $M_{200} = 10^{10.3} M_\odot$ at $z \sim 1$. This bump is not present in the X17 model prediction and is evident in the RPS-LR model only down to $z \sim 6$, while at lower redshifts this model version predicts a dip (a spoon-like feature) around $M_{200} = 10^{9.6} M_\odot$, which becomes more evident going towards $z = 0$.

At the same time we observe that, the average stellar mass of central galaxies, in bins of halo mass, predicted from the three models is similar for $M_{200} \geq 10^{10.5} M_\odot$ and continues to be similar among the X17 model and the RPS-ER model also at lower halo masses. In the case of the RPS-LR model

we observe instead that the average stellar mass of central galaxies in halos witgh $M_{200} < 10^{10.5} M_{\odot}$ is significantly higher than in the RPS-ER model, at all redshifts, and the discrepancy among the two model version increases at higher z.

What we observe for the stellar mass is motivated by the fact that in the late reionization scenario the lower values of the predicted filtering mass, based on the Kravtsov formula, allows more cold gas accretion in galaxies hosted in intermediate-low mass halos and this leads to more stars formed at high redshift in these halos. At the same time this causes the larger HI content in very low mass halos, that we see in fig. 4.2 at $z = 6.2$, while at lower z we see the effect of the fact that the predicted value of M_F, based on the Kravtsov formula, is bigger at $z < 4$ in the late-reionization scenario respect to the early-reionization one.

The observed properties regarding the average cold gas content in model galaxies reflect also into the predicted evolution of the comoving density of HI and cold gas in the universe, shown with solid and dot-dashed lines, respectively, in Fig. 4.4. We can observe that the HI cosmic density predicted by the model version RPS-ER has relatively higher values than the one predicted by RPS-LR, for $z < 4$ and the HI density predicted by the X17 model is below both the predictions by the RPS model versions, because (as shown in Xie et al., 2020b) the HI mass in galaxies predicted by the RPS model is larger than the one predicted in the X17 at all stellar masses, even if the cold gas mass is not.

What changes significantly between the X17 model and the default RPS model is the atomic-to-molecular ratio in the ISM, which assumes larger values in the RPS model (as shown in Fig.4 in Xie et al., 2020b). Looking at Fig. 4.5 we can observe that the H_2 cosmic density (dashed lines) predicted by the RPS model is a factor 3 below the predictions from the X17 model, at all redshifts, while the total cold gas cosmic density (solid lines) is associated to slightly lower estimates in the RPS model with respect to the X17 model.

In conclusion, the GAEA-RPS model predictions regarding the total comoving HI density, associated to galaxies, do not change significantly if we assume a LR scenario instead of an ER one.This result is similar to what found in the semi-analytical model developed by Croton et al. (2006) even if the treatment of the cold gas between the two models differ. The model by Croton et al. (2006) did not include an explicit treatment of the partition of

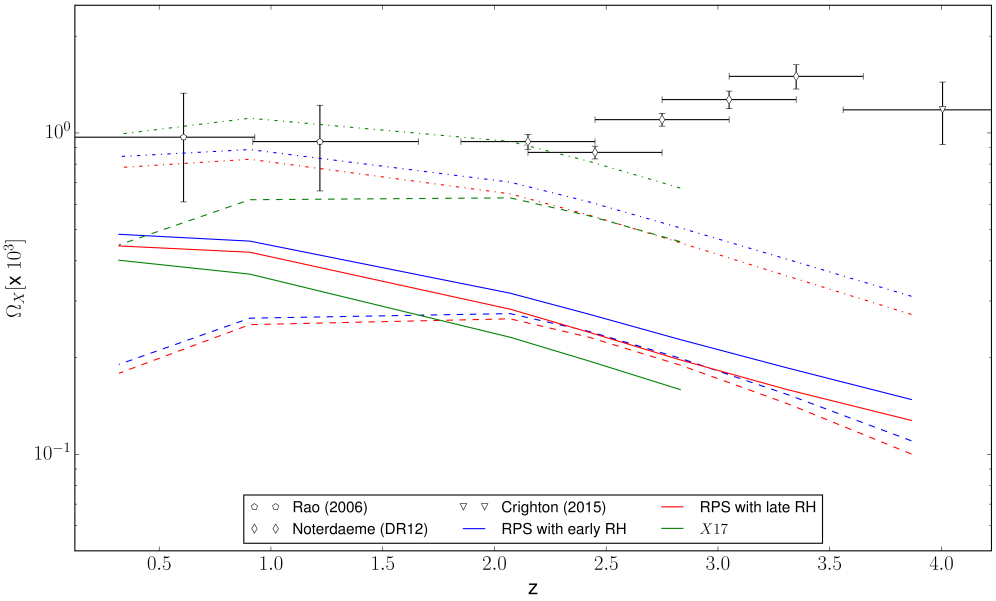

Fig. 4.5: Predicted cosmic density of HI, H_2 and CG for all the galaxies in the MSII box whith $log(M_\star) > 3$, respectively expressed with solid, dashed and dot-dashed lines, as a function of redshift. As described in the legend the predictions from the X17, RPS-ER and RPS-LR are shown in green, blue and red, respectively. The data points refer to observational constraints on the HI cosmic density, coming from DLA observations.

the cold gas in its atomic and molecular gas component, while we trace self-consistently these two components throughout the entire evolution history of the simulated galaxies.

4.3 Modifications to the prescription for the UVB feedback

Almost a decade ago, Okamoto et al. (2008b) performed high-resolution hydro-dynamical cosmological simulations with a time-dependent UV background (based on Haardt & Madau, 2001) and they found that the characteristic mass at which the halos retain only half of the typical baryon fraction is $10^7 h^{-1} M_\odot$ at z_r and $6.49 \times 10^9 h^{-1} M_\odot$ at $z = 0$, one order of magnitude lower than the one predicted by Gnedin (2000) . Their results suggested that the evolution of this characteristic mass is determined by the temperature of the accreting gas at a density of the order of $\Delta_{vir}/3$ while the filtering mass

by Gnedin (2000) was based on the average temperature of the IGM. The results by Okamoto et al. (2008b) are similar to the ones obtained by Hoeft et al. (2006), who simulated void regions.The different results obtained by Gnedin (2000) are probably due to insufficient resolution of the simulation used. Based on the results obtained by Okamoto et al. (2008b), Macciò et al. (2010) introduced a correction factor redshift-dependent in the analytic law for the filtering mass evolution derived from Kravtsov et al. (2004):

$$M_{F,new} = \frac{M_F(1+z)^{1.1}}{11.8}$$

and this is the first alternative to the implementation of the UVB feedback proposed by Gnedin (2000), that we are going to test in our new model.

The second alternative prescription that we are going to test is the result of the work done by Kim et al. (2015), inspired by the work of Sobacchi & Mesinger (2013). Sobacchi & Mesinger (2013), developed a suite of 1D simulations of collapsing shells, exploring a wide parameter space motivated by the inhomogeneity of reionization, and therefore of the UV background. With this approach they were able of exploring different scenarios of UVB feedback and deriving the following expression for the critical halo mass, below which baryons photo-evaporate or cannot efficiently accrete onto their host haloes:

$$M_c(z) = M_0 J_{21}^a \left(\frac{1+z}{10}\right)^b \left[1 - \left(\frac{1+z}{1+z_{IN}}\right)^c\right]^{d\,2} \tag{4.4}$$

with best-fit parameters

$$(M_0, a, b, c, d) = (2.8 \times 10^9 M_\odot, 0.17, -2.1, 2, 2.5)$$

The GALFORM semi-analytical model (Cole et al., 2000) is the first model which implemented the prescription by Sobacchi & Mesinger (2013) for the UVB feedback. This model reproduces the observed HI mass function atz=0, accurately matching its amplitude and shape at intermediate and high HI masses. In 2015, Kim et al. (2015) studied the effect of photoionization feedback on the HI mass function, in particular on the low mass-end of the HIMF, using the GALFORM semi-analytic galaxy formation, and assuming

[2]This formula reproduces the simulation results with a relative error of 10%, as shown in fig.2 of their article.

different parametrizations of the Sobacchi critical mass and different redshift range for Reionization. In particular Kim et al. (2015) expressed the critical mass in terms of the halo circular velocity:

$$M_c(z) = 3.36 \times 10^5 \left(\frac{V_{cut}(z)}{\mathrm{kms}^{-1}} \right)^3 (1+z)^{-3/2} M_\odot$$

where

$$V_{cut}(z) = V_{cut,0}(1+z)^{\alpha_v} \left[1 - \left(\frac{1+z}{1+z_{\mathrm{IN}}} \right)^2 \right]^{\frac{2.5}{3}}$$

$V_{cut,0}$ is the circular velocity of dark matter haloes at $z = 0$ below which the photo-ionization feedback suppresses gas cooling, and $\alpha_v = b/3 + 1/2$.

Adopting the same redshift used by Sobacchi & Mesinger (2013) for the beginning of Reionization ($z = 10$) they found that the GALFORM model gives a better match to the observed HI mass function if they instead of $V_{cut,0} = 100[km/s]$ (value corresponding to the M_0 adopted by Sobacchi) use $V_{cut,0} = 50[km/s]$. It is worth noticing that this assumption leads to a critical mass which is reduced by a factor of 10, at $z = 0$, with respect to the value advocated by Sobacchi & Mesinger (2013), as shown in fig. 4.6.

In Fig 4.6 I compare the evolution with z of all the different threshold masses coming from the different prescriptions described above, that implement the effects of the UV background on galaxies formation in low mass halos.

I expect that the effect of choosing different prescriptions for the UVB feedback on galaxies would be larger than the effect of different reionization histories because the prescriptions by Macciò et al. (2010) and Kim et al. (2015) lead to very different redshift evolution of the halo mass at which we assign only half of the baryons to galaxies.

First I analyzed the predicted redshift evolution of the average cold gas and HI content in central galaxies, shown in Fig. 4.7 and Fig. 4.8, respectively.

Both the average M_{CG} and M_{HI} associated to central galaxies in each halo mass bin grow significantly at higher redshift in the model **RPS-LR-K** , while in the model **RPS-LR-M**[3] , where I assumed also the new late reionization scenario, we can observe only a small increase at lower halo

[3] all the different models are described in table 4.1

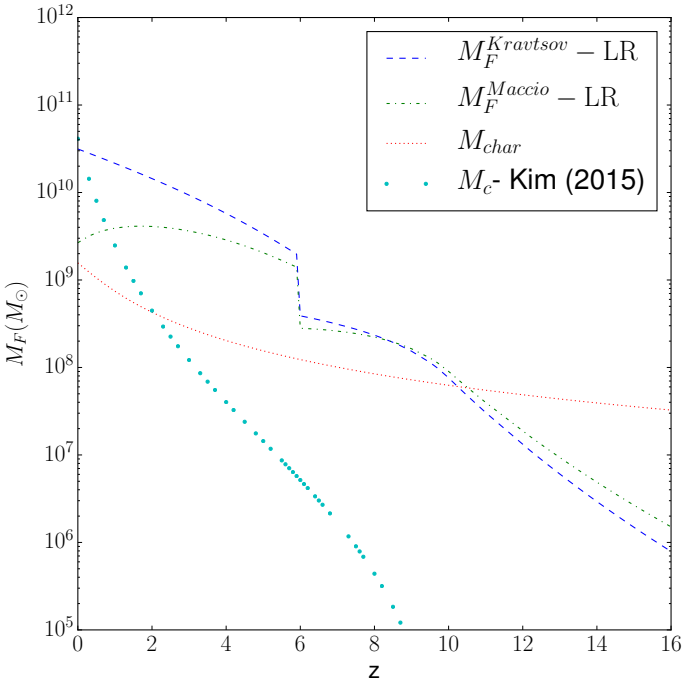

Fig. 4.6: Redshift evolution of the filtering/critical mass with different prescriptions. The label LR in the legend refers to the late reionization scenario. The Kravtsov prescription (identified by the blue line, as in Fig.4.2) and the Macciò prescription for M_F depend on the redshift range in which Reionization occurs while the prescription by Kim et al. 2015 depends only on the redshift at which Reionization started, and we adopt $z_{IN} = 10$, similarly to what is assumed in the LR scenario.

Model name	RH	UVB prescription
RPS-ER	early	Kravtsov (2004)
RPS-LR	late	Kravtsov (2004)
RPS-LR-M	late	Macciò (2010)
RPS-LR-K	late	Kim (2015)

Table 4.1: In this table there is a list of the RPS model variants which differ for the adopted reionization scenario and the adopted prescription for the UVB feedback

masses. The evolution observed in the **RPS-LR-K** model is ideal for getting in our model an evolution of $\Omega_{\rm HI}$ similar to the observed one. However, with the present calibration of the RPS model if we adopt the Kim's prescription for the UVB feedback the RPS model predicts unrealistic low values of $M_{\rm CG}$ and

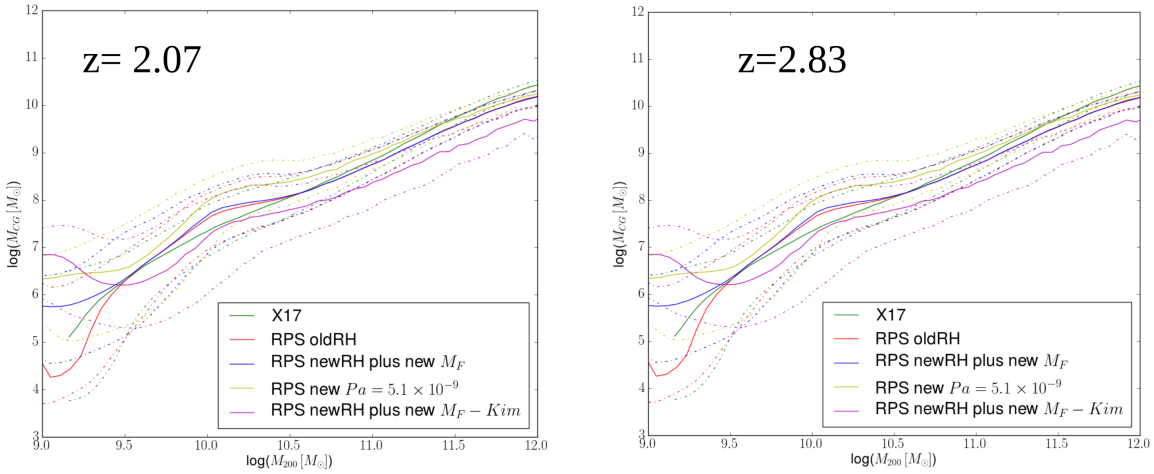

Fig. 4.7: Average CG mass for central galaxies in bins of halo mass, at $z = 2.07$ and $z = 2.83$. In the legend the different colors are associated to the X17 model, the RPS-ER Kravtsov, RPS-LR-Kravtsov, the RPS-LR-Maccio, and the RPS-LR-Kim runs, respectively. The solid lines describe the average evolution while the dot-dashed lines refer to the 1-σ range.

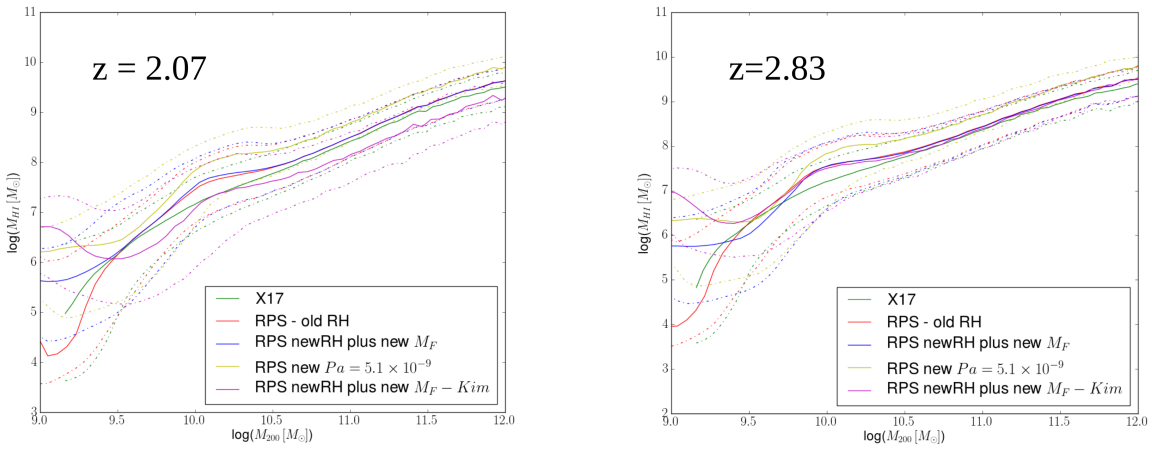

Fig. 4.8: Average HI mass for central galaxies in bins of halo mass, at $z = 2.07$ and $z = 2.83$. The line colors and styles are the same of fig. 4.7

$M_{\rm HI}$ for more massive halos at low z and it does not satisfy the observational constraints on the HIMF, coming from the local Universe, as shown in Fig. 4.9 where we can observe that the predicted HIMF (at $z = 0$) is not in agreement with the constraints coming from the HI radio surveys in the local Universe (e.g ALFALFA , HIPASS). At the same time also the predicted SMF at $z = 0$ is

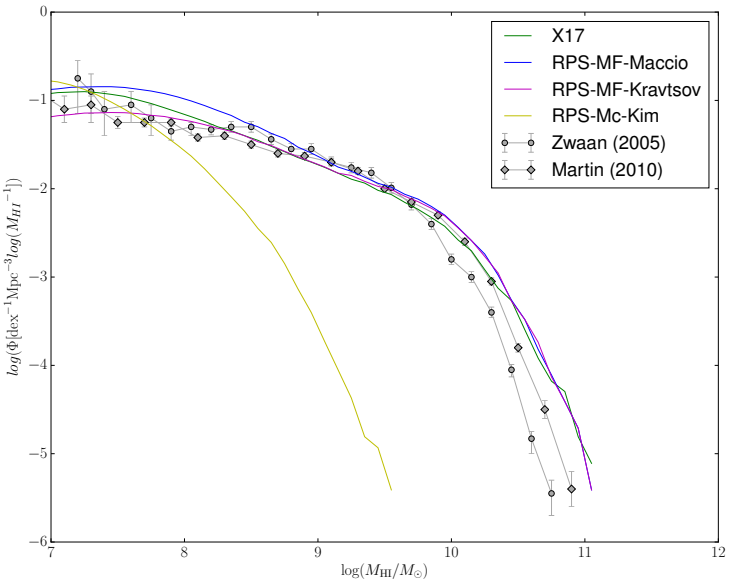

Fig. 4.9: In figure the HIMF predicted from the X17, RPS-mod-Maccio, and RPS-mod-Kim model is shown in green, light-blue and red, respectively. The grey dots and diamonds describe the observed HIMF at $z = 0$, from Zwaan et al. (2005) and Martin et al. (2010), respectively.

not in agreement with the observed one, as shown in Fig. 4.10. In this case the under-prediction of the number of objects at the high-mass end is even more severe.

It is still to verify whether it is possible to find a re-calibration of the RPS model which can reproduce the observations in the Local Universe, without losing this desirable evolution of the $\Omega_{\rm HI}$ with z.

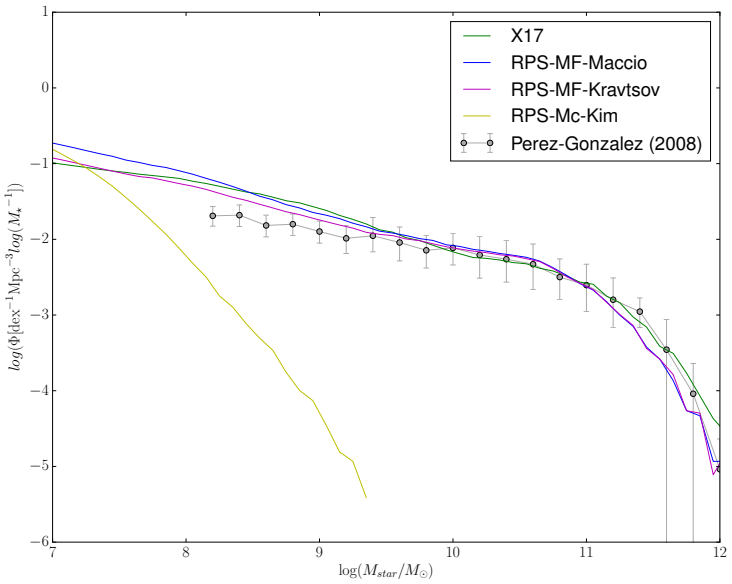

Fig. 4.10: In figure are shown the predictions from the X17, RPS-mod-Maccio, and RPS-mod-Kim model in green, light-blue and red, respectively. What is actually shown in figure is a proxy of the SMF, based on a subsample of 125 galaxy merger trees from MSII, but in the final version of the thesis there will be the SMF at $z = 0$.

We take, therefore, as new reference prescription for the UVB feedback in the RPS model, the prescription by Macciò et al. (2010). This prescription produces a better distribution of the HI in halos of different masses, in agreement with the observed HIMF at $z = 0$, as shown in Fig. 4.9.

Then I tested how the predicted CDDF in the redshift range $2 < z < 3$ changes assuming in one case the prescription for M_F by Kravtsov et al. (2004) and in the other one that by Macciò et al. (2010). To derive the predicted CDDF in the reshift range of interest I adopt the GAEA-RPS model coupled to the Millennium simulation suite in combination with a basic HOD model (described in Chapter 3). In particular I extract DM haloes with M_{200} in the mass range $(10^{9.2}, 10^{11.5})M_{\odot}$ from the MSII box, more massive haloes from the MSI box, and less massive haloes from the HOD boxes. The selected mass ranges are motivated by the resolution and volume of the two N-body simulations adopted (for more details see Sect.2 in Chapter 3). As shown in Fig 4.11 the CDDF predicted by the GAEA-LR-M model does not differ very much from the one predicted by the GAEA-LR model. In particular, we can observe that adopting the prescription by Macciò et al. (2010) for M_F we get an increase at low column densities. This small, but not negligible, increase is due to an increase in the relative contribution of intermediate/low mass haloes to the HI density at these redshifts, as shown in Fig. (4.12). Even if the predicted CDDF obtained in the LR scenario with the Maccio prescription has a shape that is more in agreement with observations, this new implementation does not solve the problem of the normalization offset present between the simulated CDDF and $\Omega_{\rm DLA}$ and the observed ones, because the total cold gas density in our simulated Universe is still too low respect to the observed one.

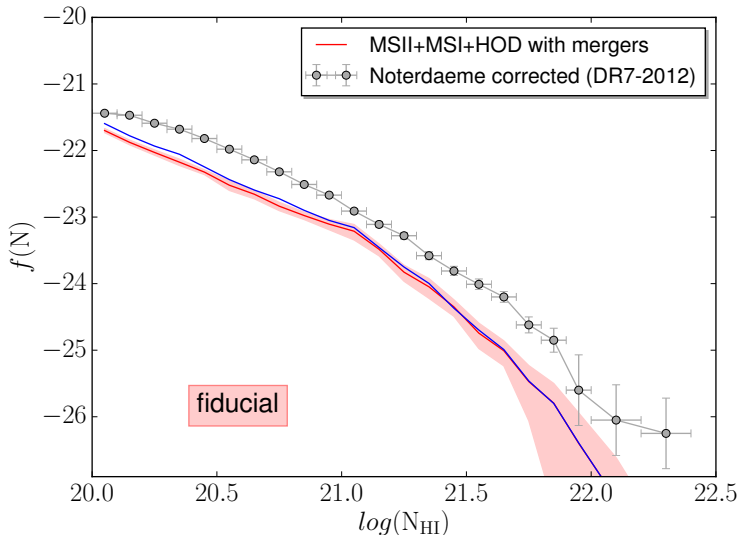

Fig. 4.11: Average CDDF in the redshift range $2 < z < 3$ predicted from the RPS model, with late-reionization scenario. The model prediction based on the prescription for M_F by Macciò et al. (2010) is shown in blue and the prescription by Kravtsov et al. (2004) in red. The shaded area represent the 68% range around the median of the predicted CDDF in the redshift range $2 < z < 3$.

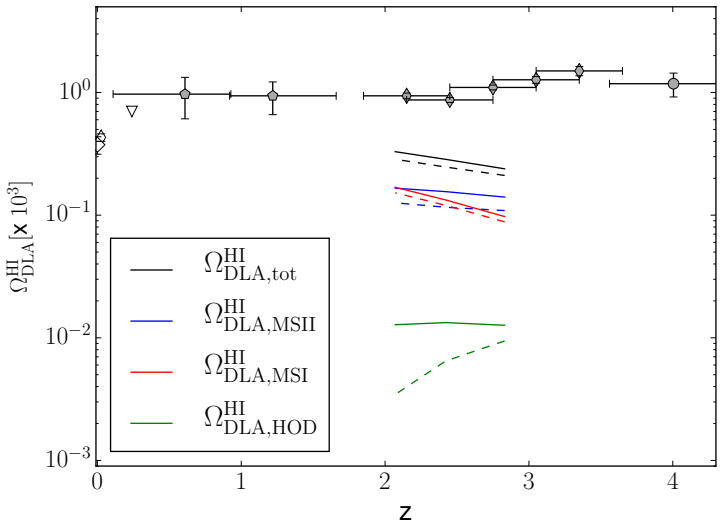

Fig. 4.12: Predicted cosmic density of HI in DLA hosted in halos in intermediate mass range (MSII), high-mass range (MSI) and low-mass range (HOD model). The color associated to different lines highlight the partial contribution to $\Omega_{\rm DLA}$ of halos in selected mass ranges, as described in the legend. The solid lines describe the predictions of the **RPS-LR** model while the dashed lines the ones of the **RPS-LR-M** model

4.4 Testing different cold gas partitioning schemes

In addition to the tests and improvements described in the previous sections I have also investigated the effect of assuming different recipes for the cold gas partitioning on the stellar and cold gas content of galaxies in the redshift range $2 < z < 3$.

In the last decade the mechanisms governing interstellar HI-to-H2 conversion and the R_{mol} profile have become an important issue in the study of galaxy evolution. For our study this subject is very important since Changes in the HI content of simulated galaxies reflect in changes of the amount of HI associated to simulated DLAs. Galaxy surveys suggest that on global scales the star-formation efficiencies are determined, at least in part, by molecular gas fractions that may be sensitive to the varying mid-plane gas pressures and/or metallicities, both at low (Bigiel et al., 2008, 2011; Fu et al., 2010; Leroy et al., 2008b, 2013; Lagos et al., 2011; Bolatto et al., 2013) and high redshift (e.g., Hirashita & Ferrara 2005; Genzel et al. 2010; Daddi et al. 2010; Tacconi et al. 2013, Popping et al. 2014)

It is generally assumed that H_2 formation enhances low-temperature cooling and cloud fragmentation or that the molecular formation rates are elevated in the denser and more shielded components of the gravitationally collapsing regions. In both cases the atomic to molecular conversion is a for having star formation in galaxies. Globally, the transition to H2 appears to be associated with star-formation thresholds and with the observed critical gas mass surface densities above which star-formation becomes probable. In spiral galaxies, the transition between a mostly–HI and a mostly–H2 ISM is a well–defined function of local conditions, and can be linked to characteristic radii, stellar and gas surface densities. In dwarf galaxies the observations reveal a more complex picture. The observations of disk galaxies on large scales (e.g. Leroy et al., 2008b) and individual Galactic molecular clouds on small scales (e.g., Lee et al. 2012,2014), indicate that for solar metallicity the HI-to-H2 conversion occurs for characteristic gas surface densities of $\sim 10 M_{\odot} \mathrm{pc}^{-2}$. This surface density corresponds to an FUV dust optical depth ~ 1, for typical grain properties and dust-to-gas mass ratios, suggesting that dust absorption and hence metallicity is playing an essential role in setting the critical gas surface densities.

For the moment, I focused my attention on three of the cold gas parti-

tioning schemes already implemented in the GAEA model: the BR scheme (Blitz & Rosolowsky, 2006), the KMT-09 scheme (Krumholz et al., 2009b), and the GK-11 scheme (Gnedin & Kravtsov, 2011). The BR scheme is an obvious choice[4] since it is the most used scheme in galaxy evolution studies to derive the cold gas partitioning in the ISM, and it has been shown to produce a better agreement with the constraints coming from local Universe (Xie et al., 2017). In addition, I have chosen to investigate the predictions of the GAEA-RPS model with the KMT and GK11 schemes for the cold gas partitioning because these two schemes are associated to a smoother decline at high redshift of the predicted $\rho_{\rm HI}$, leading to an higher value of $\rho_{\rm HI}$ at $z > 1$, as shown in Fig. 4.13 (Fig. 13 of Xie et al. (2017)), possibly reducing the discrepancy that we have at $z = 2$ between the predicted $\Omega_{\rm DLA}$ and the observed one.

However, it is worth noticing that in the last years some new models for the cold gas partitioning, based on non planar radiative transfer calculations, have been developed, to take into account the self-shielding and the dust attenuation of the local IR field (Gnedin & Draine, 2014; Sternberg et al., 2014). Sternberg et al. (2014) extended the analytic treatment for the HI-to-H2 transition presented by Sternberg et al. 1988 (S88). Their work present a detailed overview and discussion of the basic theoretical ingredients and parameters controlling the HI-to-H2 transition in FUV irradiated clouds: the FUV radiation intensity and dust attenuation cross section, the cloud gas density, and the metallicity-dependent H2 formation rate coefficient. According to the model by Sternberg et al. (2014) the HI-to-H2 transition profiles and the total atomic column densities are controlled by a single dimensionless parameter, "αG", first introduced by S88. In a nutshell, αG determines the LW-band optical depth in the cloud due to the dust associated with the HI gas (which we refer to in this paper as "HI-dust"), and whether or not the HI is mixed with the H2. In general, "αG" is the dimensionless "free parameter" in the problem, with a value that is determined by local conditions (density, radiation intensity, metallicity, etc.). However, as described by KMT, on global scales in star-forming galaxy disks the gas density and radiation inten-

[4] In our previous work, where I analysed the properties of simulated DLAs predicted by the X17 run of the GAEA model, I considered only the BR prescription for the cold gas partitioning, because Xie et al. (2017) showed that adopting different schemes for the cold gas partitioning does not affect significantly the properties of simulated galaxies predicted by the X17 model.

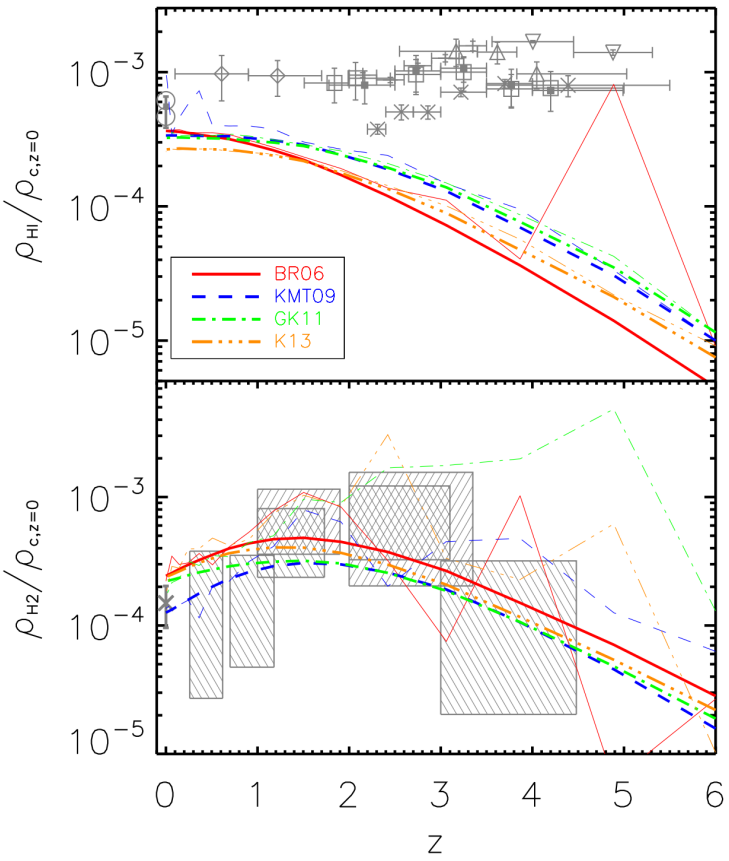

Fig. 4.13: The top panel shows the cosmic density evolution of HI while the bottom panel shows the H_2 cosmic density. Different colours and line styles correspond to the different star formation laws considered in Xie et al. (2017) work, as indicated in the legend. Thick lines correspond to densities estimated considering all galaxies down to the completeness limit of the MSII. Thin lines have been obtained by fitting the H I and H 2 mass functions at different redshifts and extrapolating them towards lower masses.

sity may be correlated or self-regulated to conditions enabling a two-phase equilibrium between cold and warm HI (CNM/WNM multiphase). This then drives "αG" to a narrow range of values, near unity, and the predicted HI column density becomes only weakly dependent on the metallicity (in a way similar to the KMT prediction)

In the following I will describe the prescriptions adopted in each scheme for the derivation of the molecular gas fraction ($R_{\rm mol}$) and the physical motivations at the basis of each scheme. Then I will describe the galaxy properties resulting from the adoption of each particular scheme (in particular SMF and

HIMF). Finally I will move to the predicted DLA properties and I will discuss the results.

4.4.1 The BR model

The BR prescription is based on the relation observed in local galaxies between the ratio of molecular to atomic hydrogen ($R_{\rm mol}$) and the mid-plane pressure ($P_{\rm ext}$) acting on the galactic disc (Blitz & Rosolowsky, 2006). The formula for the ratio of molecular to atomic hydrogen derived by Blitz & Rosolowsky (2006) is:

$$R_{\rm mol,BR} == \frac{\Sigma_{\rm H_2}}{\Sigma_{\rm HI}} = \left(\frac{P_{\rm ext}}{P_0}\right)^{\alpha} \tag{4.5}$$

where P_0 is the external pressure of molecular clumps. Based on their sample of 14 nearby galaxies, Blitz & Rosolowsky (2006) find P_0 ranging between $0.4 \times 10^4 {\rm cm}^{-3}$ K and $7.1 \times 10^4 {\rm cm}^{-3}$ K, and values for α varying between 0.58 and 1.64. We assume $\log(P_0/k_B[{\rm cm}^{-3}{\rm K}]) = 4.54$ and $\alpha = 0.92$, that correspond to the mean values. The hydro-static pressure at the mid-plane can be written as follows (Elmegreen, 1993):

$$P_{\rm ext} = \frac{\pi}{2} G \Sigma_{\rm gas} [\Sigma_{\rm gas} + f_\sigma \Sigma_\star] \tag{4.6}$$

where $\Sigma_{\rm gas}$ and $\Sigma_\star$ are the surface density of the cold gas and of the stars, respectively, and they are estimated in each one of the 20 annulii for each modeled galaxy (as described in detail in Xie et al., 2017), and $f(\sigma) = \sigma_{\rm gas}/\sigma_\star$ is the ratio between the vertical velocity dispersion of the gas and that of the stellar disk.

In the GAEA model we assume a constant velocity dispersion for the gaseous disk, equal to $\sigma_{\rm gas} = 10\,{\rm km\,s^{-1}}$ (Leroy et al., 2008a), while for the stellar disk we follow Lagos et al. (2011) and assume $\sigma_\star = \sqrt{\pi G h_\star \Sigma_\star}$ and $h_\star = r_{\star,{\rm d}}/7.3$, based on observations of nearby disc galaxies (Kregel et al., 2002).

Following Lagos et al. (2011), we assume for this model:

$$\nu_{\rm sf,br} = \nu_{\rm br,0} [1 + \left(\frac{\Sigma_{\rm gas}}{\Sigma_{0,\rm br}}\right)^{0.4}] \tag{4.7}$$

where $\Sigma_{0,\rm br} = 200\,{\rm M}_\odot\,{\rm pc}^{-2}$ is the critical density where the slope of the relation between $\Sigma_{\rm SFR}$ and $\Sigma_{\rm H_2}$ steepens (Bigiel et al., 2008). $\nu_{\rm br,0} = 0.4\,{\rm Gyr}^{-1}$ is chosen to reproduce the observed H_2 mass function and galaxy stellar mass function at z=0.

4.4.2 The KMT09 model

In a series of papers (Krumholz et al., 2008, 2009a,b) Krumholz et his collaborators developed an analytic model to determine the fraction of molecular hydrogen, within a single atomic-molecular complex, resulting from the balance between dissociation of molecules driven by interstellar radiation, self-shielding of the molecular clouds, and formation of molecules on the surface of dust grains.

In this model the molecular gas fraction is described by Eq. 4.8:

$$f_{\rm mol,kmt} = \frac{\Sigma_{\rm H_2}}{\Sigma_{\rm H_2} + \Sigma_{\rm HI}} = 1 - [1 + \left(\frac{3}{4}\frac{s_{\rm kmt}}{1+\delta}\right)^{-5}]^{-1/5}, \tag{4.8}$$

where,

$$s_{\rm kmt} = \ln(1 + 0.6\chi_{\rm kmt})/(0.04\,\Sigma_{\rm comp,0}\,Z'), \tag{4.9}$$

$$\chi_{\rm kmt} = 0.77(1 + 3.1\,Z'^{0.365}), \tag{4.10}$$

$$\delta = 0.0712\,(0.1 s_{\rm kmt}^{-1} + 0.675)^{-2.8}, \tag{4.11}$$

and

$$\Sigma_{\rm comp,0} = \Sigma_{\rm comp}/(1\,{\rm M}_\odot{\rm pc}^{-2}). \tag{4.12}$$

$\Sigma_{\rm comp}$ is the surface density of a giant molecular cloud (GMC) on a scale of ~ 100 pc, and Z' is the metallicity of the gas normalized to the solar value (we assume $Z_\odot = 0.02$). Following Krumholz et al. (2009b), we assume $\Sigma_{\rm comp} = f_c \Sigma_{\rm gas}$, where f_c is a 'clumping factor' that assumes the value of 1 on scales close to 100 pc, and that we treat as a free parameter of the model. In previous studies, the values assumed for this parameter range from 1.5 (Fu et al., 2010) to 5 (Lagos et al., 2011). In our case, $f_c = 3$ provides predictions that are in reasonable agreement with data, while larger values tend to under-predict the HI content of massive galaxies. Krumholz et al. (2009b) stress that some of the assumptions made in their model break at gas metallicities below roughly 5 per cent solar ($Z' < 0.05$). As discussed e.g. in Somerville et al. (2015), we can assume that the gas in the galaxies

resolved in our model is already enriched to metallicities $\sim 10^{-3} Z_{\odot}$ at high redshift, since POP III stars and other enrichment mechanisms were active at very high redshift and in the mass regime that our model cannot resolve. Following their approach, when computing the molecular fraction with the KMT recipe, we assume this floor in case the metallicity of the cold gas would result to be lower.

4.4.3 The GK11 model

Gnedin & Kravtsov (2011) carried out a series of high resolution hydro-simulations including non-equilibrium chemistry and an on-the-fly treatment for radiative transfer, in order to follow the formation and photo-dissociation of molecular hydrogen, and self-shielding in a self-consistent way. Gnedin & Kravtsov (2011) provide a fitting function that describes the fraction of molecular hydrogen as a function of the dust-to-gas ratio in units of the dust-to-gas ratio of the Milky Way ($D_{\rm MW}$), the intensity of the radiation field (G'_0), and the gas surface density ($\Sigma_{\rm gas} = \Sigma_{\rm HI+H_2}$). The fitting formula is reported in Eq. 4.13 :

$$f_{\rm mol,gk} = \frac{\Sigma_{\rm H_2}}{\Sigma_{\rm gas}} = [1 + \frac{\Sigma_{\rm c}}{\Sigma_{\rm gas}}]^{-2}, \tag{4.13}$$

where $\Sigma_{\rm c}$ is a characteristic surface density of neutral gas at which star formation becomes inefficient.

$$\Sigma_c = 20\,{\rm M_{\odot}pc^{-2}} \frac{\Lambda^{4/7}}{D_{\rm MW}} \frac{1}{\sqrt{1 + G'_0 D^2_{\rm MW}}}, \tag{4.14}$$

with:

$$\Lambda = \ln(1 + g D^{3/7}_{\rm MW} (G'_0/15)^{4/7}), \tag{4.15}$$

$$g = \frac{1 + \alpha_{\rm gk} s_{\rm gk} + s^2_{\rm gk}}{1 + s_{\rm gk}}, \tag{4.16}$$

$$s_{\rm gk} = \frac{0.04}{D_{\star} + D_{\rm MW}}, \tag{4.17}$$

$$\alpha_{\rm gk} = 5 \frac{G'_0/2}{1 + (G'_0/2)^2}, \tag{4.18}$$

$$D_{\star} = 1.5 \times 10^{-3} \ln(1 + (3G'_0)^{1.7}), \tag{4.19}$$

Following GK11, we use the metallicity of cold gas to get the dust ratio: $D_{\rm MW} \approx Z' = Z_{\rm gas}/Z_{\odot}$. For G'_0, it is assumed

Within the simulation run Gnedin & Kravtsov (2011) varyed $D_{\rm MW}$ from 10^{-3} to 3, and G'_0 from 0.1 and 100. In the GAEA model it is assumed $D_{\rm MW} = 10^{-3}$ to calculate the molecular fraction when the cold gas metallicity $Z_{\rm gas} < 10^{-3} Z_{\odot}$.

GK11 also provide the star formation efficiency necessary to fit the observational results in Bigiel et al. (2008) in their simulations:

$$\nu_{\rm sf,gk} = \frac{1}{0.8\,{\rm Gyr}} \times \begin{cases} 1 & \Sigma_{\rm gas} \geq \Sigma_{\rm gk} \\ (\frac{\Sigma_{\rm gas}}{\Sigma_{\rm gk}})^{\beta_{\rm gk}-1} & \Sigma_{\rm gas} < \Sigma_{\rm gk} \end{cases} \tag{4.20}$$

where $\Sigma_{\rm gas}$ is the surface density of cold gas, $\Sigma_{\rm gk} = 200\,{\rm M}_{\odot}{\rm pc}^{-2}$, and $\beta_{\rm gk} = 1.5$.

4.4.4 A modified BR scheme

Since the BR scheme is based on observational constraints of galaxy properties in the Local Universe, it is possible that the parameters adopted in this prescription would change at higher redshift, where the ISM is much more turbulent, the local UV field is stronger and the gas phase metallicity in galaxies on average is lower. In the last decade astrophysicists learned a lot about the ISM in high-z galaxies (e.g. cold gas fraction, SFR, ionization level), thanks to precise measurements of ISM properties at high-z (ALMA,MUSE), and high-resolution simulations, e.g. FLARES (Lovell et al., 2021), FIRE-2 (Ma et al., 2018; Ma, 2018; Ma et al., 2020), CROC (Zhu et al., 2020), SPHINX (Rosdahl et al., 2018), RENAISSANCE (Barrow et al., 2017), BLUETIDES (Wilkins et al., 2017). However, most of the studies (in particular the observational ones) are focused on individual galaxies and we still lack a comprehensive and coherent picture that links together the formation of molecular gas, star formation, the turbulent, multi-phase structure of the ISM, and the importance of the various feedback channels due to star formation.

In 2011 Feldmann et al. developed a suite of high-resolution hydrodynamical simulations of galaxies in a cosmological framework, to study the origin of the $P_{ext} - R_{mol}$ relation and its dependence on the properties of the ISM (Feldmann et al., 2011). In particular, they analysed the properties of the cold ISM as a function of the dust-to-gas (DTG) ratio and the strength of

the local UV field. Their simulation showed that, under MW-like ISM conditions, a relation similar to BR arises naturally, and also galaxies with different DTG ratios and/or radiation fields follow a $P_{ext} - R_{mol}$ relation, but with changes in the normalization and the slope. In particular, they showed that the amplitude of the relation $R_{mol} - P_{ext}$ (linked to the value of $1/P_0$) correlates with the dust-to-gas ratio measured in the ISM, which is expressed by means of $[O/H]$.

This theoretical result together with the observed decrease of the average DTG in neutral gas absorbers (tracers of galaxies) with increasing z, as shown in Fig.11 in the review by Péroux & Howk (2020) , implies that assuming that the value of P_0 increases towards higher z is physically motivated.

In particular, I implemented a prescription for the cold gas partitioning in the ISM of modelled galaxies, based on the BR scheme, where I vary $P_a = A/P_0$ as described in eq. 4.20 :

$$P_a(z) = \frac{P_a(0)}{(1 + 4.5z)} \tag{4.21}$$

The redshift evolution of P_a described in the equation above reflects in the reference value of P_0 at $z = 0$ and in the value $P_0 = 4.7 \times 10^{-13}$Pa at $z = 2$ (1 order of magnitude higher respect to the value at $z = 0$).

I will refer to the run where I applied this modified BR scheme, as **RPS-mod-BR-evo-P_0**.

In parallel I have also tested the effect of assuming a different value for the exponent α ($\alpha = 0.8$, instead of $\alpha = 0.92$) in the BR law, which describes the dependence of R_{mol} on $\mathrm{P_{ext}}$. I will refer to this run as **RPS-mod-BR-new-α**.

4.4.5 Results

The first things that I have tested are the HIMF and the SMF predicted by the RPS model versions based on five different schemes for the cold gas partitioning in the ISM of model galaxies, illustrated in previous subsections. In order to have a lower computational cost, for the analysis of the effects of different CG partitioning schemes on modeled galaxies, I have considered only the galaxies inside the MSII box. In Fig. 4.14 and in Fig. 4.15 are shown the predicted HIMF and SMF, respectively, at three different redshifts ($z = (0.3, 0.95, 2.07)$), together with relative observational constraints.

We can observe that the **RPS-mod-GK** and the **RPS-mod-KMT** runs predict a bump in the HIMF at intermediate HI masses ($10^8 M_\odot < M_{\rm HI} < 10^{10} M_\odot$) at $z = 0$, and this bump is not in agreement with observational constraints on the HIMF (Zwaan et al., 2005; Martin et al., 2010). The bump is present at all redshift, but it becomes flatter and covers a narrower mass range at higher redshift.

The other model versions predict an HIMF at $z = 0.0$ on average in agreement with observations. In particular the X17 model and the **RPS-mod-BR-new** α model give very similar predictions, while the **RPS-mod-BR-evo**P_0 model predicts a bump in the mass range $10^8 M_\odot < M_{\rm HI} < 10^{9.2} M_\odot$, which becomes more evident at higher redshift, and an higher number density of galaxies at the high-mass end of the HIMF.

To understand the origin of the bump in the HIMF predicted by the **GAEA-RPS-GK** model and not observed in the **GAEA-RPS-BR** model run, I have investigated the histories of 500 galaxies, selected to have HI mass values in the range covered by the bump in the HIMF at $z = 0$ and binned in halo mass. Comparing the evolution histories predicted for these galaxies by the **GAEA-RPS-GK** models with the ones predicted by the **GAEA-RPS-mod-BR-evo**P_0 we observe that the SFR associated to these galaxies in the **GAEA-RPS-mod-BR-evo**P_0 model run is constantly one order of magnitude (or more) higher respect to the other run, for $z < 1$, and this causes a more rapid growth of the ejected mass and stellar mass, and a decrease of the cold gas mass towards $z = 0$.

Moving to the analysis of the predicted redshift evolution for the SMF, we can observe that the X17 model, the **RPS-mod-BR-evo**P_0 and the **RPS-mod-BR-new** α runs predict very similar SMF at $10^9 M_\odot < M_\star < 10^{11} M_\odot$. At the same time, the **RPS-mod-GK** and the **RPS-mod-KMT** runs strongly under-predict the number density of galaxies at the high-mass end of the SMF and over-predict the number density in the mass range $10^8 M_\odot < M_\star < 10^9 M_\odot$, creating an excess at low masses in the predicted SMF, which increases at higher redshifts. The differences in the predicted slope for the very high-mass end of the SMF between the BR and the KMT/GK runs, are due to the larger disks associated to high-mass galaxies in the GK and KMT model runs with respect to what observed in the BR run. However, since in this mass regime we have very few galaxies in the MSII box, due to the little volume ($L_{box} = 100 {\rm cMpc}$), we need to check the model predictions for the SMF at

$M_\star > 10^{12} M_\odot$ running the model on the cosmological box associated to the MSI.

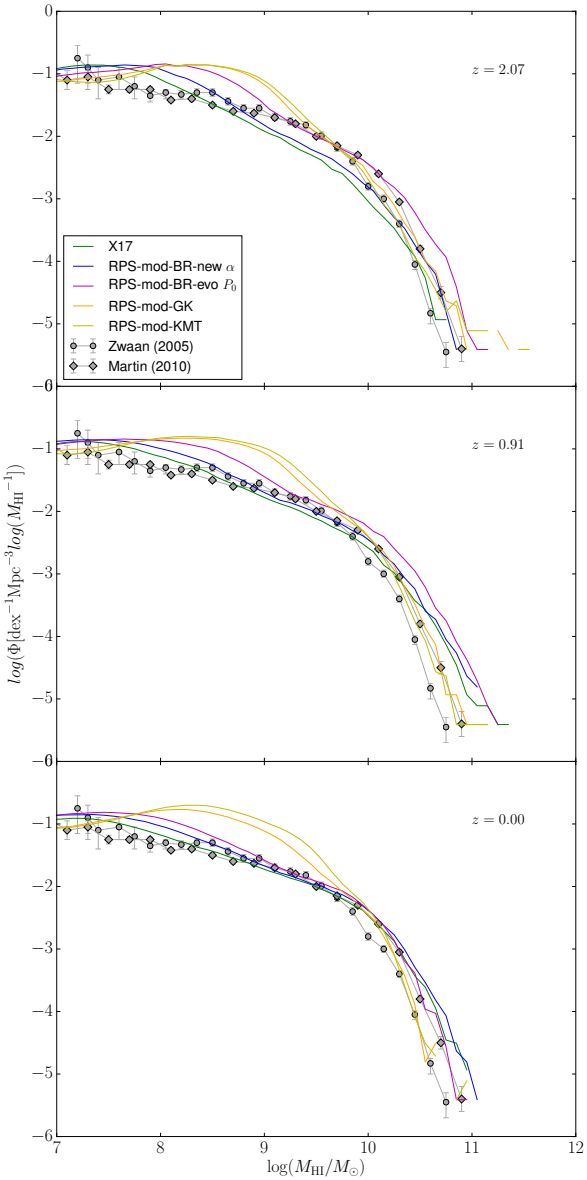

Fig. 4.14: HIMF predicted from different GAEA model versions for all the galaxies within the MSII box, at three different redshifts, from $z = 0.3$ to $z = 2.07$. The data points refer to the observed HIMF at $z = 0$ Zwaan et al. (2005); Martin et al. (2010).

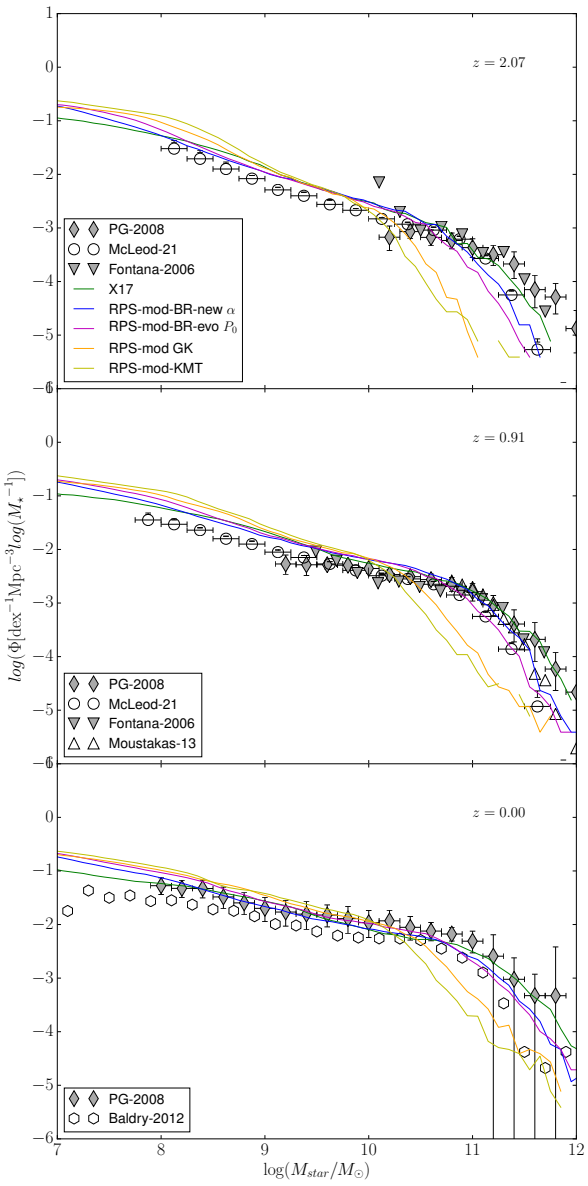

Fig. 4.15: SMF predicted from different GAEA model versions for all the galaxies in the MSII box, at three different redshifts, from $z = 0.0$ to $z = 2.07$. The data points refer to the SMF constrained at different z by different galaxy surveys with multi-band observations (Pérez-González et al., 2008; Baldry et al., 2012; Moustakas et al., 2013; Fontana et al., 2006; McLeod et al., 2021)

Then I looked at the contribution to the predicted CDDF, in the redshift range $2 < z < 3$, coming from the haloes well resolved in the MSII run, and I investigated how this quantity changes with the adopted partitioning scheme. Since we have seen that the prediction from the GK model do not differ too much from the KMT ones I focused only on the RPS-mod-BR-evo-P_0 model, the RPS-mod-α model, and the RPS-mod-KMT model, and I compared the predictions of these models to the ones of the RPS-mod-BR model.
In Fig. 4.16 it is shown the predicted contribution to the DLA CDDF of different halo mass bins, well resolved in the MSII simulation, for different **RPS-mod** runs adopting different scheme for the cold gas partitioning in the ISM, compared to the observed total CDDF in the redshift range $2 < z < 3$ (Noterdaeme et al., 2012). Different line styles in the figure indicate different runs of the RPS model:

- solid lines $\rightarrow$ **RPS-mod-BR**: RPS model with late reionization scenario combined to the prescription of M_F from Macciò et al. (2010) and BR partitioning scheme with $P_a = 5.1 \times 10^{-8}$, $\alpha = 0.92$
- dashed lines $\rightarrow$ **RPS-mod-BR-evo**P_0: RPS model with late reionization scenario+ prescription of M_F from Macciò et al. (2010) and BR partitioning scheme with evolving P_0, $\alpha = 0.92$
- dot-dashed lines $\rightarrow$ **RPS-mod-BR** α: RPS model with late reionization scenario + prescription of M_F from Macciò et al. (2010) and BR partitioning scheme with $P_a = 5.1 \times 10^{-8}$, $\alpha = 0.8$
- dotted lines $\rightarrow$ **RPS-mod-KMT**: RPS model with late RH + prescription of M_F from Macciò et al. (2010) and KMT partitioning scheme

We can notice that the **RPS-mod-KMT** run is the GAEA-RPS run that predicts the highest contribution to the DLA CDDF in all the halo mass bins contributing, but it requires a strong recalibration of the model to get predictions in agreement with the observational constraints from the local Universe. The **RPS-mod-BR-evo**P_0 run predicts a slightly lower estimate of the contribution to DLA CDDF coming from the halo mass bins selected, but its predictions are still in agreement with the contribution needed to reproduce the observed DLA CDDF at this redshift: their values are similar to the ones obtained with the model $2M - 2R$ described in Chapter 3 (see fig. 3.6) The predicted contribution to the DLA CDDF in the **RPS-mod-BR** α run

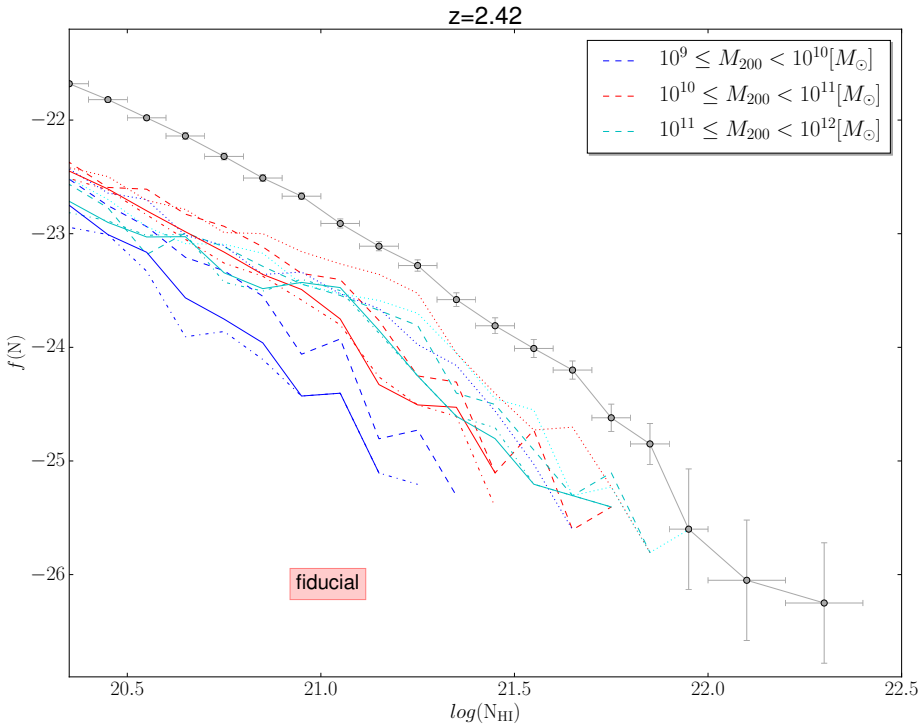

Fig. 4.16: Average contribution to the CDDF at $z = 2.44$ of galaxies in different halo mass bins, resolved in the MSII bins, predicted from the RPS-mod model coupled to different schemes for the cold gas partitioning. Data points refer to the observed total CDDF in the redshift range $2 < z < 3$ (Noterdaeme et al., 2012)

is even lower than in the **RPS-mod-BR** run, which assumes the traditional BR scheme.

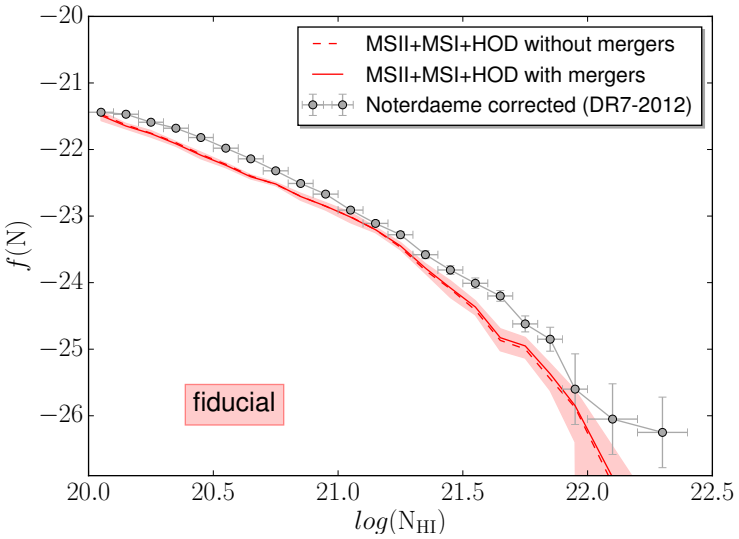

Fig. 4.17: Average predicted CDDF for $2 < z < 3$ from the **RPS-mod-BR-evo** P_0 model. Data points refer to the observed total CDDF in the redshift range $2 < z < 3$ (Noterdaeme et al., 2012).

At this point I focused my attention on the **RPS-mod-BR-evo** P_0 model, and I estimated the predicted CDDF for $2 < z < 3$, combining the MSI,MSII and HOF halo catalogs as described in the previous section (and in Sect. 3.2, in more detail). The predicted CDDF is shown in Fig. 4.17.
It is on average a bit offset low with respect to the observed CDDF (Noterdaeme et al., 2012).

I have also fitted the Γ function to the CDDF from the **RPS-mod-BR-evo** P_0 and the **RPS-mod-BR** model, in the redshift range $2 < z < 3$. The best-fit parameters are in Table 4.2. Comparing the value of the turnover column density (N_g) associated to the CDDF predicted by the two RPS model version with the values found by Zafar et al. (2013) and Noterdaeme et al. (2012) in a similar redshift range, we can observe that our value is a bit higher than the one derived by Zafar et al. (2013) ($log(N_g)$ =21.08) and more similar to the one derived by Noterdaeme et al. (2012), but still in agreement with both within the errors.

Model	α_G	$log(\kappa_g)$	$log(N_g)$
RPS-BR-evoP_0	-1.31353166	-23.08000404	21.26858604
RPS-BR	-1.41337516	-23.28749544	21.24589739

Table 4.2: Best-fit parameters of the Γ function fits to the average CDDF in the redshift range $2 < z < 3$, for different RPS models.

In parallel, I analysed the predicted $\Omega_{\rm DLA}$ for $2 < z < 3$, finding what is shown in Fig.4.15. We have decided to derive estimates of $\Omega_{\rm DLA}$ only in the redshift range of interest for this study. Indeed, the agreement between the **RPS-mod-BR-evo**P_0 model predictions and observations on the HI content of galaxies (and therefore DLAs) at $z = 0$ is shown in Fig. 4.14 (where I plot the predicted HIMF).

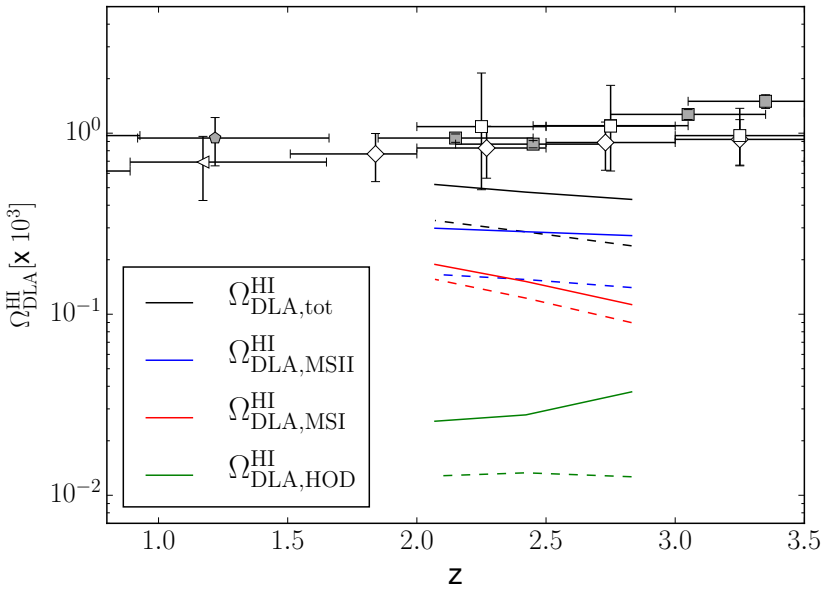

Fig. 4.18: Predicted cosmic density of HI in DLA hosted in halos in intermediate mass range (MSII),high-mass range (MSI) and low-mass range (HOD model). The color associated to different lines are described in the legend. The solid lines are associated to the model **RPS-mod-BR-evo** P_0 while the dashed lines to the **RPS-mod-BR**. The grey data points refer to observational constraints on the HI cosmic density, coming from DLA observations, for the cosmology adopted in Crighton et al. (2015), while the white points refer to more recent observational estimates and are taken from the review by Péroux & Howk (2020)(see references therein).

The total comoving density of HI in simulated DLAs predicted by this model version is slowly declining with redshift and higher then half the value of the observed $\Omega_{\rm DLA}$. In the redshift range of interest the major contribution to $\Omega_{\rm DLA}$ comes from the haloes in the mass range covered by the MSII, whose contribution changes more in the model **RPS-mod-BR-evo**P_0 with respect to the model **RPS-mod-BR**.
In conclusion, adopting a late reionization scenario and the prescription for

the filtering mass by Macciò et al. (2010), the GAEA variant that reproduces in a better way the observed DLA statistics in the redshift range $2 < z < 3$, without losing the agreement between the model predictions and the observational constraints on the HIMF and SMF coming from the local Universe, is the **RPS-mod-BR-evo-**P_0 model.
We will consider hereafter the **RPS-mod-BR-evo**P_0 run our reference model. And we will test the model predictions regarding other important galaxy e DLA properties: metallicity and H_2 content.

4.4.6 DLA metallicity

In this paragraph I will illustrate the properties of the metallicity distribution associated to the simulated DLAs in the redshift range of interest ($2 < z < 3$). I will show four different plots which describe different ways to estimate the metallicity associated to simulated DLAs. In the plots shown in Fig. 4.19 I adopt the same procedure described in Chapter 3, and I associate only the metal content of the cold gas in the ISM to DLAs, in the first case distributing it uniformly over the galactic disk (top panel) and in the second case adopting a radial metallicity gradient with a universal slope (bottom panel), equal for all the simulated galaxies. In the plots shown in Fig. 4.20 I associate to simulated DLAs the sum of the metal content of the cold gas in the ISM, the hot gas and the ejected gas. As previously done, I have chosen to distribute it uniformly over the galactic disk in one case (top panel) and in the other I adopted a radial metallicity gradient with a universal slope (bottom panel), equal to the slope estimated by Christensen et al. (2014a)for all the simulated galaxies.The results shown in Fig. 4.19 demonstrate that considering only the metals in the cold gaseous phase of the ISM to derive the DLA metallicity, the contour plots of the distribution of the predicted DLA metallicities are not in agreement with the distribution of the observed metallicities, also for the new GAEA model (**GAEA-RPS-mod-BR-evo**P_0), similarly to what we found for GAEA-X17 (see Chapter 3). Instead, considering also the metals in the ejected component associated to the DLA host galaxy to derive the DLA metallicity, we are capable to reproduce better the observed bidimensional distribution of $[Fe/H]$ VS $N_{\rm HI}$. Even if, it is important noticing that our model still predicts the existence of some metal-poor DLAs, under the observed DLA metallicity floor.

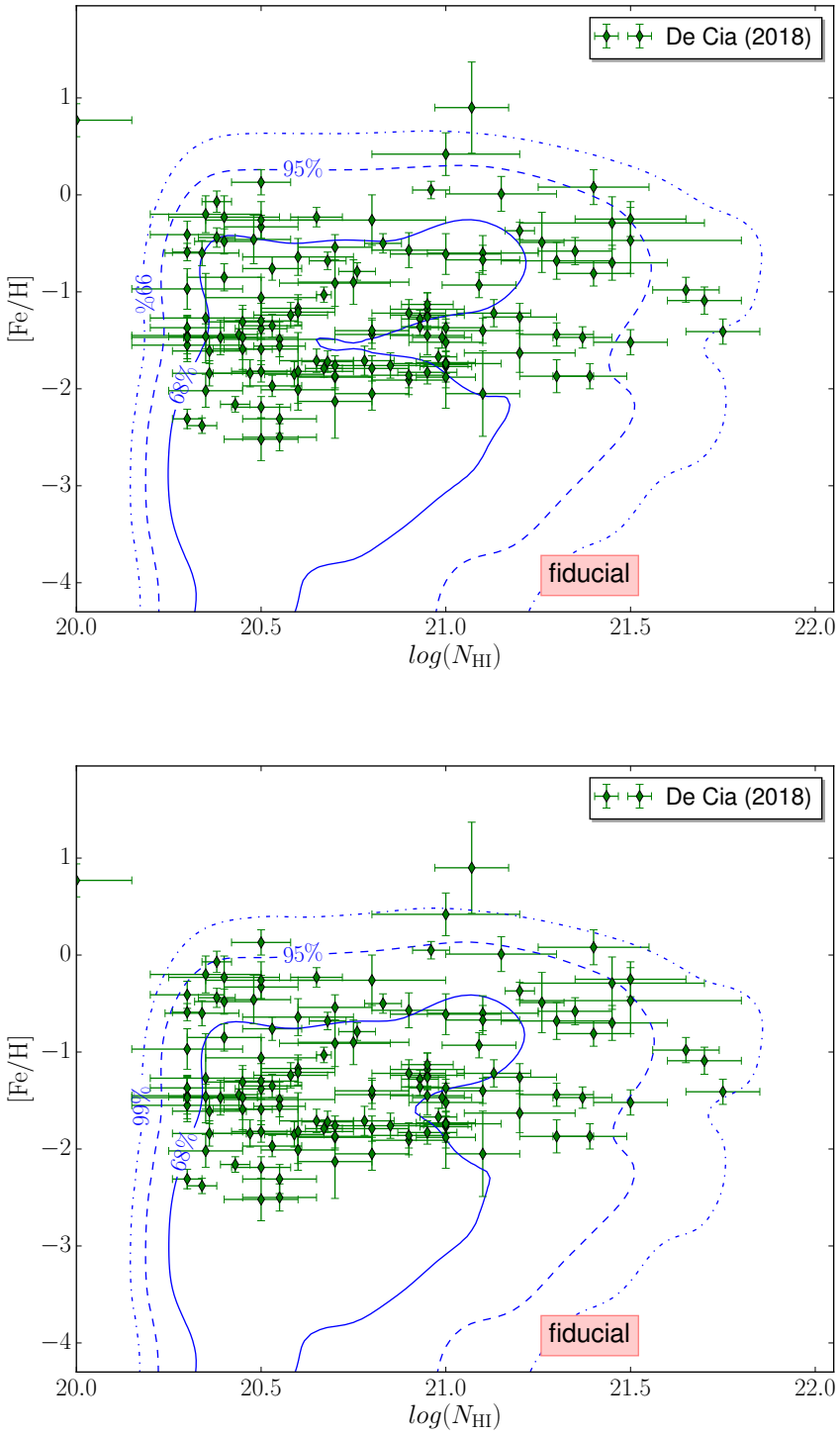

Fig. 4.19: $[Fe/H]$ VS $N_{\rm HI}$ for simulated DLA in the reshift range $2 < z < 3$. In this plot the abundance ratio describes only the metal content associated to cold gas in the ISM.

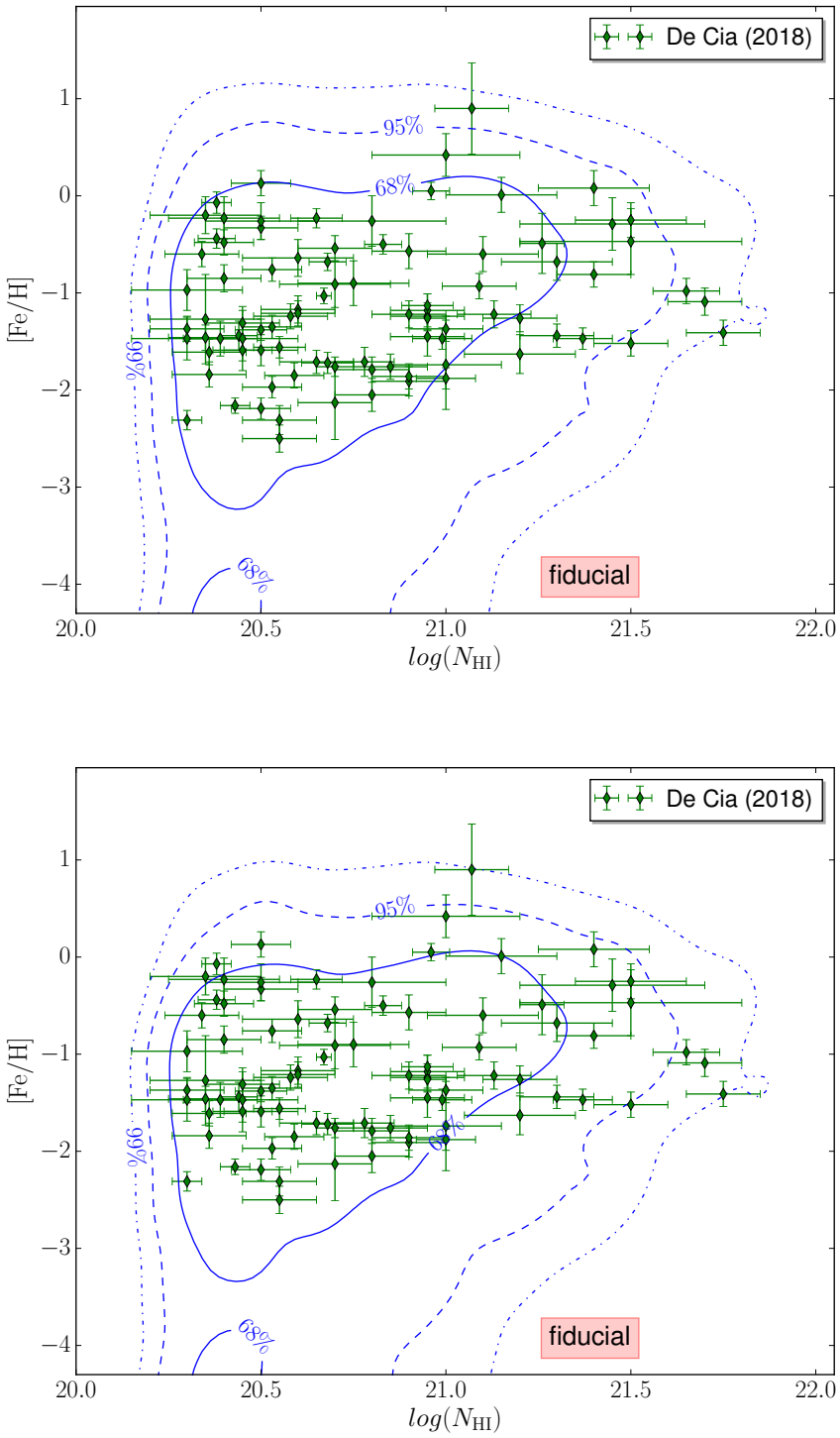

Fig. 4.20: $[Fe/H]$ VS $N_{\rm HI}$ for simulated DLA in the reshift range $2 < z < 3$. In this plot the abundance ratio describes the metal content associated to cold gas in the ISM, hot gas and ejected gas.

4.5 How do model predictions compare with the observational constraints on the molecular content in DLAs and galaxies?

Until now we have explored the modifications in the HI content of model galaxies induced by different prescriptions for the cold gas partitioning and reionization, neglecting the effects on the molecular gas content. To have a more complete picture of how the different prescriptions affect the galaxy evolution in the modeled Universe, it is important to compare the predictions regarding the molecular content in galaxies, with the available observations at all redshifts.

The last decade has brought impressive developments in surveys of molecular gas, and in particular molecular hydrogen (H_2), in resolved and unresolved galaxies, locally and at high redshift (e.g. Daddi et al., 2010; Genzel et al., 2010; Davis et al., 2011; Saintonge et al., 2011; Young et al., 2011; Boselli et al., 2014; Decarli et al., 2016, 2019).

Molecular hydrogen is very difficult to observe in the ISM of galaxies. Lacking a dipole moment its emission is extremely weak at the typical temperature of the cold ISM. The molecular content in high-z galaxies is not probed directly in emission but indirectly, with different tracers (e.g. CO lines, the CI fine structure lines, PAHs). A widely used tracer of H_2 is the carbon monoxide (CO) molecule, which is the second most abundant molecule in the ISM, and is easily excited (see Carilli & Walter 2013 for a review). The CO luminosity function is well probed in the local Universe, by means of the ground CO rotational transition CO(1-0) (Keres et al., 2003). From this luminosity function,and adopting a Milky Way (MW) like CO(1-0)$-H_2$ conversion factor (Bolatto et al., 2013), it has been derived the H_2 mass function and the cosmic density of the molecular hydrogen , ρ_{H_2}, at $z < 0.1$ (Keres et al., 2003; Obreschkow & Rawlings, 2009). Campaigns to obtain constraints on ρ_{H_2} at higher redshift have used blind CO surveys (e.g. Aravena et al., 2012, 2016; Decarli et al., 2016, 2019). In particular the ALMA Spectroscopic Survey in the Hubble Ultra Deep Field (ASPECS) covers the reshift range $0.2 < z < 4.0$ while the VLA (Very Large Array) based CO Luminosity Density at High-z survey (COLDz) made complementary measurements at $z > 4$ (Riechers et al., 2019, 2020), whereas the IRAM (Institut de Radioastronomie Millimétrique) 30-m-based xCOLD GASS (Extended CO Legacy Database for GASS-GALEX

Arecibo Sloan Survey) provided robust measurements at $z = 0$ (Saintonge et al., 2017). The estimates of the molecular hydrogen content in galaxies derived by CO observations depend strongly on the assumed conversion factor ($\alpha_{\rm CO}$). While many works have investigated the relation between $\alpha_{\rm CO}$ and the galactic environment in the Local Universe (Genzel et al., 2012, 2015; Accurso et al., 2017), it is still unclear how this conversion factor varies with galactic properties at high redshift, in particular at $z > 2$. Therefore, for the estimation of the molecular density at $z \sim 0$ Saintonge et al. (2017) adopt a $\alpha_{\rm CO}$ which accounts for the observed galaxy properties (in particular for the sSFR and metallicity), following the prescription by Accurso et al. (2017). Instead, Decarli et al. (2016, 2019) and Riechers et al. (2019) in their work adopt an universal conversion factor ($\alpha_{\rm CO}$) in the derivation of the molecular density estimates at high redshift. This leads to a strong uncertainty in the H_2 estimates at intermediate/high redshift, and to some ambiguity in the comparison between theoretical predictions, by models and simulations, and the latest results from ASPECS/COLDz surveys.
Klitsch et al. (2019) used the ALMA calibrator archive to search for CO absorption, providing stringent limits on the molecular gas density at $z < 1.5$. Deriving estimates of the molecular gas cosmic density from blind surveys for CO absorbers is promising. This technique allows to probe the lower gas column density regime (not sampled in emission line surveys), free from cosmic variance issues (see Kanekar et al., 2014, , for more details), while cosmic variance affects in a not negligible way the estimates derived by the ASPECS surveys due to the small sky area covered by this survey (Keenan et al., 2020). Some recent works showed that there is a tension between most of the up-to-date model predictions regarding the H_2 content and the results of the ASPECS survey based on the following assumption for the conversion factor: $\alpha_{\rm CO} = 3.6 M_{\odot}/(\text{K km/s } pc^2)$ (Popping et al., 2019b; Davé et al., 2019).

In particular, Popping et al. (2019b) show that when a conversion factor $\alpha_{\rm CO} = 3.6 M_{\odot}/(\text{K km/s } pc^2)$ is assumed and the ASPECS sensitivity limits[5]

[5]According to the ASPECS CO sensitivity based selection criterion, ASPECS is sensitive to sources with an $M_{\rm H_2} \sim 10^9 M_{\odot}$ at $z = 0.29$ and $\sim 10^{10} M_{\odot}$, $2 \times 10^{10} M_{\odot}$, $3 \times 10^{10} M_{\odot}$, at $z \sim 1.43$, 2.61 and 3.8, respectively (see Boogaard et al., 2019; Decarli et al., 2019, , for more details), therefore the selection effect becomes stronger at higher z (as we can observe from the comparison between the dashed and the solid lines in fig. 7 by Popping et al. (2019b)). Apart from this mass limit in their work Popping et al. (2019b) considered also the aperture effect, due to the physical extent of the beam aperture (which covers $3.5''$ in the case of the

are taken into account, both IllustrisTNG (Weinberger et al., 2017; Pillepich et al., 2018) and the Santa Cruz semi-analytic model (SC SAM Somerville & Primack, 1999; Somerville et al., 2001; Somerville et al., 2008; Somerville et al., 2015; Yung et al., 2019) predict H_2 masses that are too low at a given stellar mass for galaxies at $z > 1$, the number density of H_2 -rich galaxies (with $H_2 > 3 \times 10^{10} M_\odot$) is underestimated (fig. 4.21), and the predicted evolution for the cosmic density of H_2 is marginally compatible (SC SAM) or in tension (TNG) with the ASPECS results (fig. 4.22).

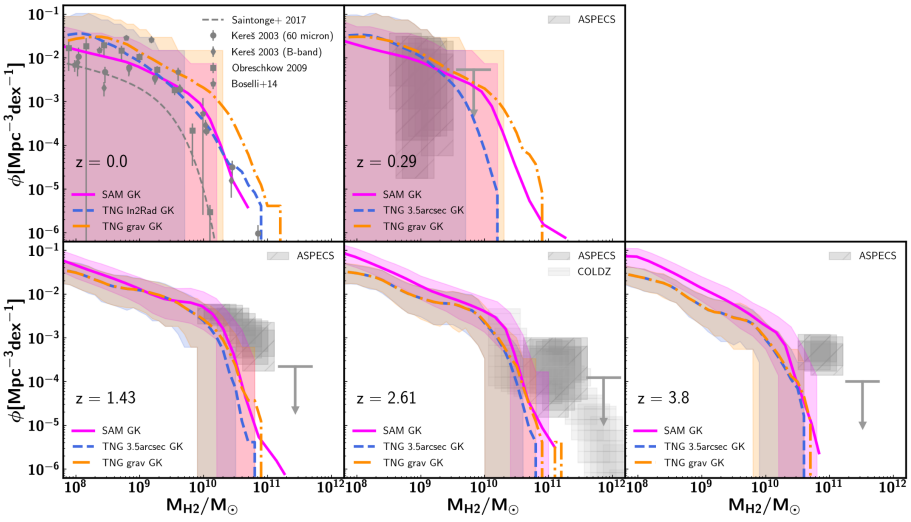

Fig. 4.21: The predicted and observed $H2$ mass function of galaxies assuming $\alpha_{\rm CO} = 3.6 M_\odot/(Kkm/spc^2)$ at all the redshifts probed by ASPECS. Model predictions are shown for the SC SAM (solid pink) and IllustrisTNG ("3.5arcsec" aperture: dashed blue; "Grav" aperture: dashed-dotted orange), both models adopting the GK partitioning recipe. In this Figure the thick lines mark the mass function based on the entire simulated box ($\sim$ 100 cMpc on a side for IllustrisTNG, $\sim$ 142 cMpc on a side for the SC SAM). The colored shaded regions mark the 2σ scatter when calculating the $H2$ mass function in 1000 randomly selected cones that capture a volume corresponding to the volume probed by ASPECS at the given redshifts. At $z = 0$ the model predictions are compared to observations from Keres et al. (2003); Obreschkow & Rawlings (2009); Boselli et al. (2014); Saintonge et al. (2017). At higher redshifts the model predictions are compared to observations from the ASPECS (Decarli et al., 2016, 2019) and COLDZ (Riechers et al., 2019) surveys

ASPECS survey), which covers only portions of the larger galaxies observed, in particular at low redshift.

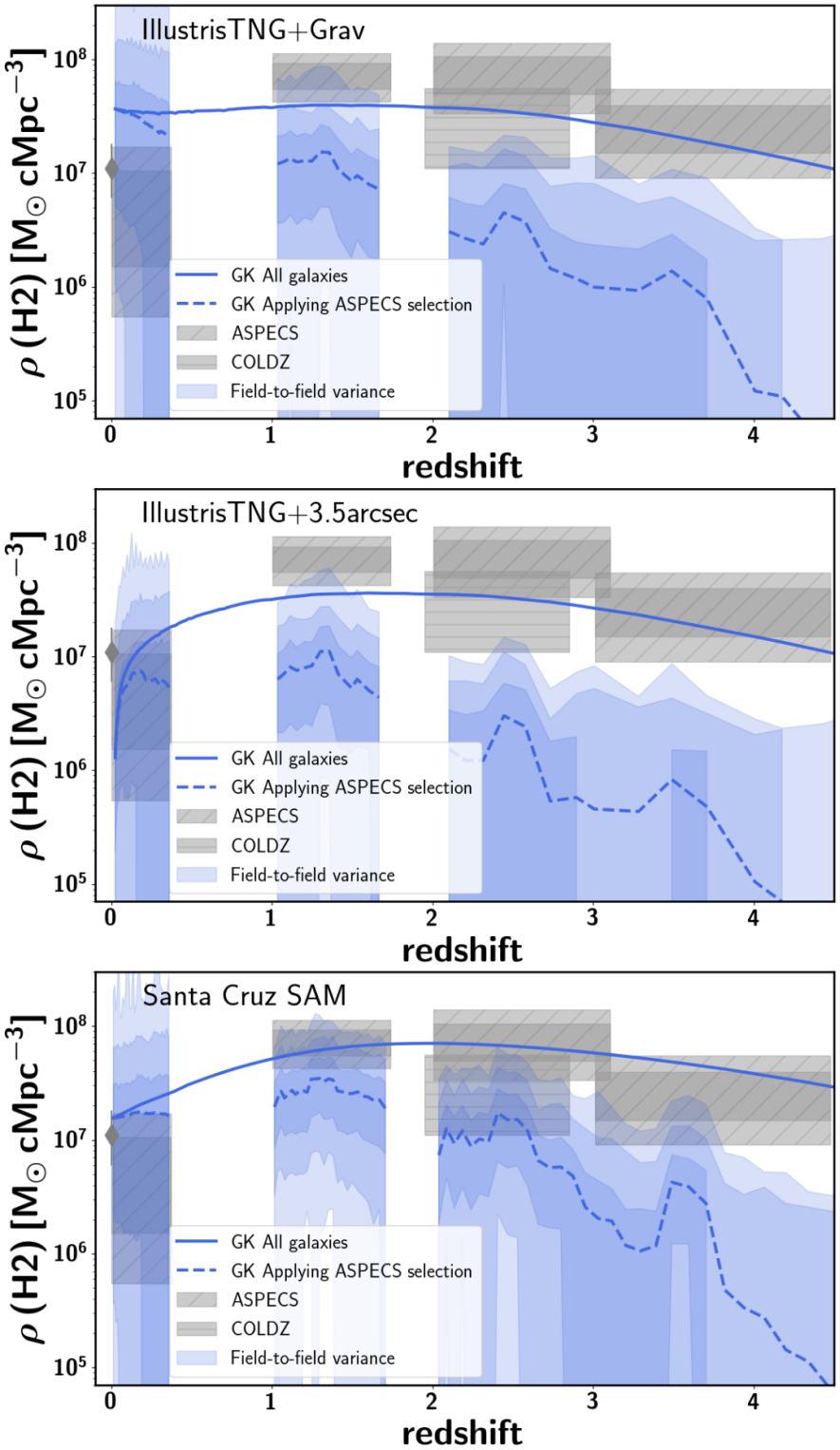

Fig. 4.22: The predicted and observed H_2 cosmic density assuming $\alpha_{\rm CO} = 3.6 M_{\odot}/(Kkm/spc^2)$, as a function of z, from IllustrisTNG ('Grav' aperture, top; '3.5arcsec' aperture, center), and the SC SAM (bottom), adopting the GK partitioning recipe. Solid lines correspond to the cosmic H_2 density based on all the galaxies in the entire simulated volume. Dashed lines correspond to derived applying the ASPECS selection function. Shaded regions mark the 0th and 100th percentiles, 2σ, and 1σ scatter when calculating the H_2 cosmic density in 1000 randomly selected cones with a volume representing the ASPECS survey. Observations are from ASPECS (dark (light) grey mark the 1(2) σ uncertainty), COLDZ (Riechers et al., 2019), and from Saintonge et al. (2017) at $z = 0$.

As shown in Fig. 4.23(fig. 6 from Popping et al. (2019b)) the H2MF predicted by the SC SAM at $1 < z < 2$ comes into better agreement with ASPECS results when adopting $\alpha_{\rm CO} = 2.0 M_{\odot}/(K$ km/s $pc^2)$ instead that the standard value of the conversion factor, but the agreement is still poor at higher z. Adopting a lower value for the conversion factor ($\alpha_{\rm CO} = 0.8 M_{\odot}/(Kkm/spc^2)$) they obtain a better agreement also at higher z, but this comes at the cost of predicting H_2 masses too high with respect to the observed ones at z = 0.

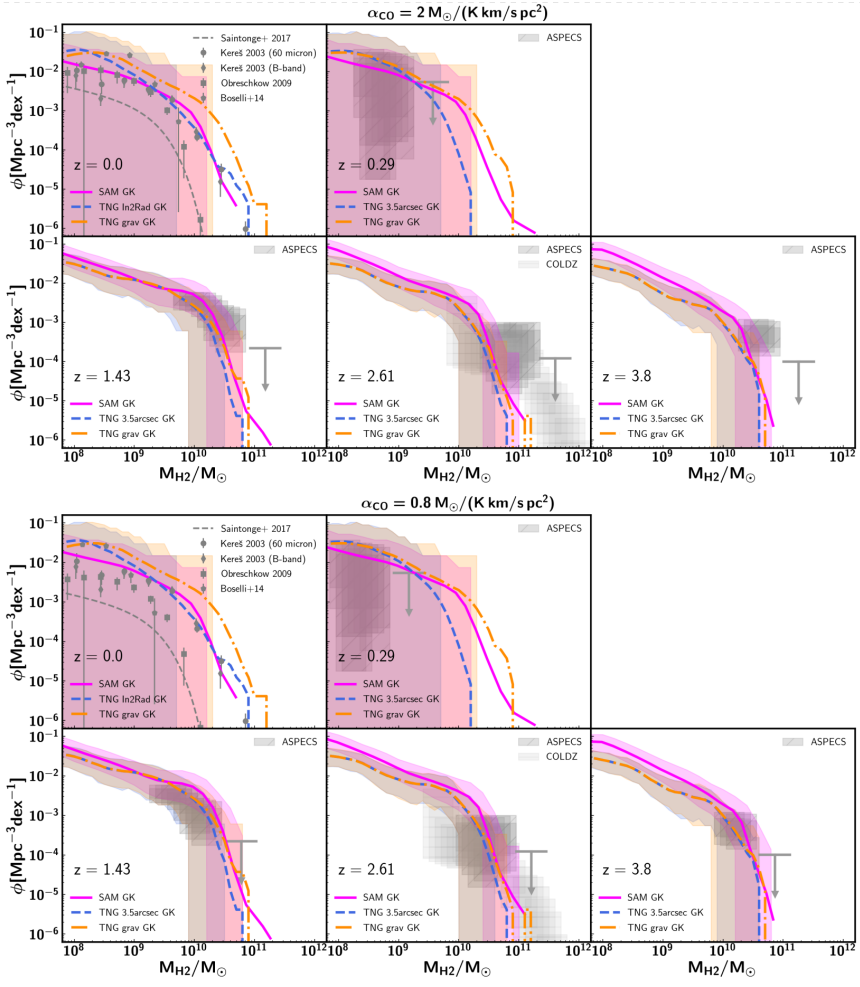

Fig. 4.23: The predicted and observed H 2 mass function of galaxies at z = 0 and the redshifts probed by ASPECS as predicted from IllustrisTNG and the SC SAM. The top five and bottom five panels correspond to a scenario where we adopt a CO-to$-H_2$ conversion factor of $\alpha_{\rm CO} = 2 M_{\odot}/$(K km/s pc^2) and $\alpha_{\rm CO} = 0.8 M_{\odot}/$(K km/s pc^2) for the observations, respectively. For details on the thick lines and shaded areas see the caption of fig. 4.16

The use of the conversion factor value $\alpha_{\rm CO} = 2.0 M_{\odot}/(K$ km/s $pc^2)$ is motivated from observations of main-sequence galaxies at $z = 1 - 3$ (e.g. Genzel et al., 2012; Popping et al., 2017). The conversion factor value $\alpha_{\rm CO} = 0.8 M_{\odot}/(Kkm/spc^2)$, typical of ULIRGs, seems unrealistic for the entire sample, but it could be justified for the brightest CO sources in the ASPECS survey.

Following Popping et al. (2019b) before comparing our model predictions for the galactic H_2 content with ASPECS data I applied the selection function[6] to the modeled galaxy sample. In particular I considered the limit in CO luminosity of the ASPECS surveys at different redshifts, which reflects into a mass limit for the molecular hydrogen content of the observed galaxies.

The first constraints on H_2 against which I tested our model predictions are the estimates, at different z, of the comoving H_2 density (ρ_{H2}). In Fig. 4.24 it is shown the comparison between the predictions of the **GAEA-RPS-mod-BR-evo**P_0 model (blue), the Santa-Cruz SAM predictions (red) and the more recent estimates on the molecular cosmic density at different redshifts. The data point refer to: the ASPECS survey (Decarli et al., 2016, 2019), the COLDz survey (Riechers et al., 2019), the PHIBSS2 survey (Lenkić et al., 2020), the COLDGASS survey (Saintonge et al., 2017), and the observation estimates by Andreani et al. (2020) and Fletcher et al. (2021) in the Local Universe. The estimates derived by Decarli et al. (2019); Riechers et al. (2020) and Lenkić et al. (2020) are based on an universal value for the conversion factor from CO-to-H2 while the estimates at $z \sim 0$ adopts a variable $\alpha_{\rm CO}$, which depends on galaxy properties. As done in Popping et al. (2019b) I show with solid lines the total cosmic H_2 density and with dashed lines the H_2 density associated to the galaxies observable with ASPECS, once applied the conversion factor from CO-to-H2. I considered three different values for the conversion factor: in the top panel I show the results for the same conversion factor adopted in the ASPECS ufficial papers, while in the center and bottom panel I adopt ($\alpha_{CO} = (0.8, 2.0) M_{\odot} [K\, kms^{-1} pc^{-2}]^{-1}$), respectively, following Popping et al. (2019b) . A different assumption for the conversion factor value implies different H_2 mass limits for the ASPECS survey and the other surveys at different redshifts and therefore different estimates for the predicted values of ρ_{H2}, as shown in Fig 4.24. I have rescaled for the differ-

[6] It takes into account the selection effects present in the survey.

ent adopted $\alpha_{\rm CO}$ value only the data points by Decarli et al. (2016, 2019); Riechers et al. (2019) and Lenkić et al. (2020). If we take a look to the comparison of total $H2$ cosmic density predicted by the two different semi-analytical model, we can observe that the SC model predict higher estimates with respect to the GAEA model, at $z > 0.5$, and the discrepancy in prediction increases at higher z. If we consider the results shown in the top panel, once considered the ASPECS selection function for observable galaxies, and assuming as conversion factor, neither of the two model can reproduce the observed evolution of ρ_{H2}, and in particular the GAEA model predicted valued that are more than an order of magnitude offset low with respect to observations. If we assume instead $\alpha_{CO} = 2.0 M_{\odot}[K\ kms^{-1}pc^{-2}]^{-1}$, the predictions of the SC model turn into agreement with the observations, while the GAEA model predicts still ρ_{H2} estimates which are offset low, as shown in the center panel. Considering instead $\alpha_{CO} = 0.8 M_{\odot}[K\ kms^{-1}pc^{-2}]^{-1}$, the SC model slightly ovepredicts the cosmic $H2$ content at all z, while the GAEA model comes into agreement with observations for $z > 1$, as shown in the bottom panel of Fig. 4.24, and it overpredicts a bit the $H2$ content in the Universe at $z < 1$. Moreover, comparing the prediction of our model with the ones coming from the SC model we can observe that the redshift evolution for ρ_{H2} predicted by the **GAEA-RPS-mod-BR-evo**P_0 model is flatter than the evolution predicted by the SC model, for $z < 2$, and the GAEA model overpredict the content of molecular hydrogen in the low-z Universe, if we believe to the last assumption on the conversion factor. It is worth noting that the shape of the H_2 density evolution, predicted by the **GAEA-RPS-mod-BR-evo**P_0 model, does not agree with the observed one for any of the assumed conversion factors from L_{CO} to $M(H_2)$,because we do not observe a peak in $\rho(H_2)$ around $z \sim 2$ and instead we observe a flattening in the evolution of $\rho(H_2)$ at low redshift. The origin of this 'peculiar' redshift evolution of $\rho(H_2)$ needs to be investigated in more detail, and I plan to do this in my future work.

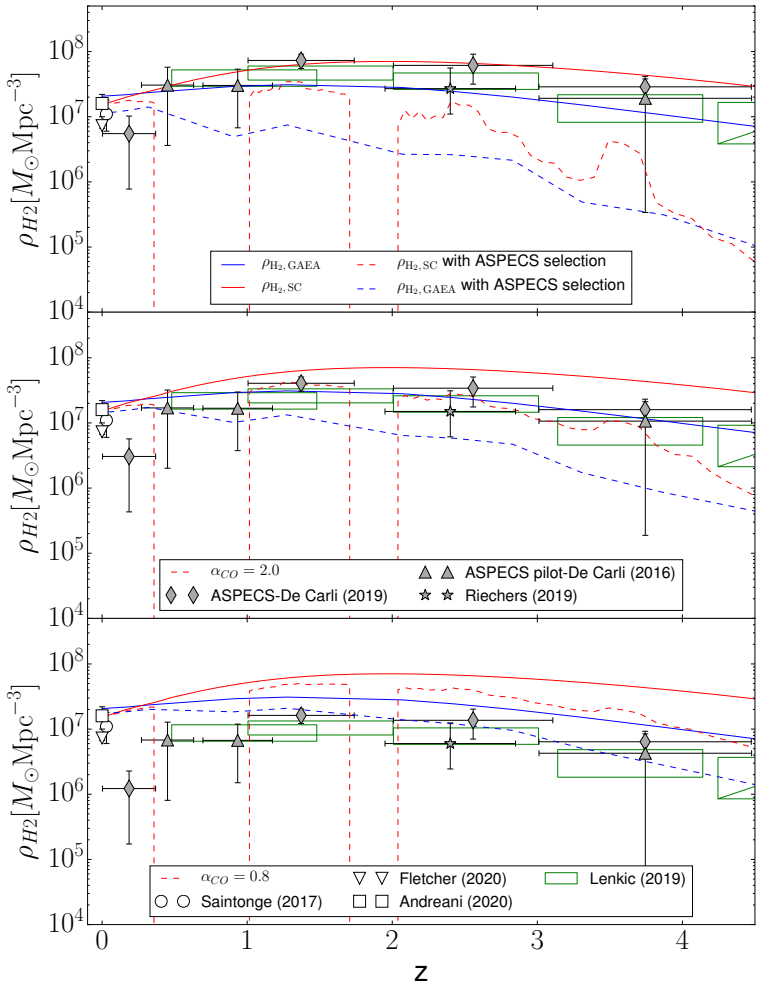

Fig. 4.24: Evolution with redshift of the cosmic density of H_2 in the **GAEA-RPS-mod-BR-evo**P_0 and SC models, compared to recent observations. The data point refers to $\rho(H_2)$ estimates derived by the works described in the legend. The values predicted by the GAEA model for Ω_{H2}, at different z, are shown in blue, while the predictions coming from the Santa-Cruz model (Popping et al., 2019b) are shown in red. The solid lines describe the evolution of ρ_{H2} without any selection effect accounted for. The dashed lines instead describe the redshift evolution of ρ_{H2} once applied the ASPECS selection, considering decreasing values for the conversion factor CO-to-H2 moving from the top panel to the bottom panel.

A first step to understand the origin of the differences in normalization and redshift evolution of the cosmic H_2 density, between the SC model and the GAEA model, is to look at the predicted H_2 mass function (H2MF) and its evolution with redshift in the two models. Therefore, we have compared the H2MF redshift evolution predicted by the GAEA model with the one predicted by the SC model.

In fig. 4.25 I show a comparison between the H2 mass function predicted by two runs of the GAEA model (GAEA-X17-BR and GAEA-RPS-mod-BR-evoP0) and the Santa Cruz model, at $z = (0.0, 0.29, 1.43, 2.61, 3.8)$. For the last 4 redshift bins considered I have added also the ASPECS H2 mass function. The first thing that we can observe is that the low mass-end of the H2MF predicted by the Santa Cruz model has a normalization higher than the one predicted by both the two GAEA runs. Moreover, the low-mass end evolution

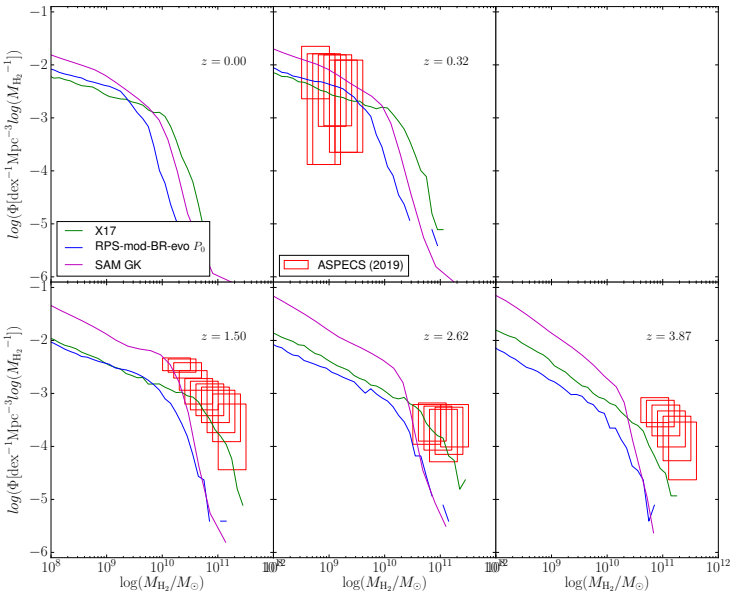

Fig. 4.25: H2 mass function at 5 different redshifsts $(0.0, 0.29, 1.43, 2.61, 3.8)$. In green and blue we show the prediction from the GAEA-X17-BR and **GAEA-RPS-mod-BR-evo**P_0 models, respectively. In magenta I show the predictions from the Santa Cruz model. The red squares describes the H2MF from ASPECS in 4 different redshift bins. It is shown only the 1-sigma box for each bin.

with redshift is different in the two models. The different evolution observed for the predicted H2MF could be explained with the different prescriptions adopted for the cold gas partitioning (it is GK for the Santa Cruz model and BR for the GAEA model), the star formation, and the stellar feedback, but also with different recipes for the cooling process. I plan to investigate the origin of this discrepancy in the near future with a detailed comparison of the predicted SMF and ρ_{SFR}, at different redshifts. At the same time the two H2MF predicted by the different GAEA model version differ mostly at the high mass end and the H2MF predicted by the X17 model seems to be more in agreement with the ASPECS data for $z > 1.2$ with respect to the one associated to the **GAEA-RPS-mod-BR-evo**P_0 model.

Moreover, I studied the column density distribution function of dense H_2 absorbers (CDDF(H_2)) associated to DLAs in the redshift range $2 < z < 3$ and in Fig. 4.26 I compared the predictions of the GAEA RPS-mod-BR-evoP0 model with the results obtained by Balashev & Noterdaeme (2018) and the upper limits derived by Klitsch et al. (2019), for $z > 0.5$, from the analysis of the results of the ALMACAL survey, a wide and deep survey utilizing the ALMA calibrator archive for the search of CO absorption lines. In their work Balashev & Noterdaeme (2018) analysed the stacked spectra obtained averaging over a large amount of quasar spectra with low S/N and intermediate resolution. They found that the covering factor of strong H_2 systems($N_{H2} > 10^{18}\mathrm{cm}^{-2}$) is about 4% in systems with $N(HI) > 10^{20} cm^{-2}$, about 9% in a metal sub-sample with prominent metal lines and about 30% in systems with $N(HI) > 10^{21.5} cm^{-2}$. The total DLA covering factor is consistent with both the results from the UVES and MagE searches for H_2 systems.

In our model we found a much higher covering factor for strong H_2 systems associated to DLAs ($\sim 85\%$ at $z = 2.5$). As we can observe in Fig. 4.26 our predicted CDDF(H_2) is almost one order of magnitude higher respect to the observed one, in all the column density bins considered. This discrepancy could be due either one or a combination of three effects:

1. we do not model properly the clumpiness of the molecular hydrogen, but we distribute it smoothly across the disk
2. an observational systematic effect which biases the results by Balashev & Noterdaeme (2018),
3. we do not consider the warm neutral phase of the ISM/CGM (where

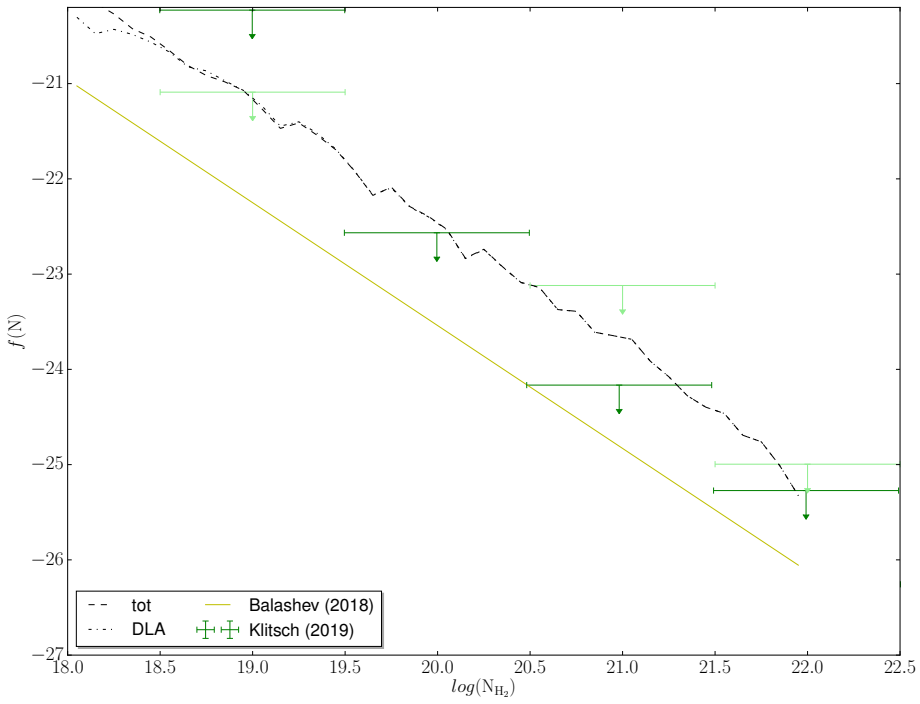

Fig. 4.26: Average predicted CDDF for H_2 associated to DLAs in the redshift interval $2 < z < 3$ from the **RPS-mod-BR-evo**P_0 model. The yellow line describe the fit to the CDDF obtained by Balashev et al. 2018, while the green data points are from Klitsch et al. (2019). The light coloured limits reflect the uncertainty introduced by the CO-to-H_2 column density conversion factor.

there is not H_2) and we associate all the simulated DLAs to the cold gas phase.

Likely, it is a combination of these three effects that causes the observed discrepancy.

In the last decade there have been tremendous efforts from the astronomical community to search for HI and OH 21 cm absorbers at $0 < z < 5$, motivated by the fact that these absorbers trace specifically the cold neutral medium (CNM) in galaxies and could tell us how the physical properties of the ISM evolve, with redshift and with the environment. This search has been challenged by the presence of RFI contamination and only recently some groups developed different technique to discriminate properly the false detections, due to RFI, from the true detections. In particular there is the ongoing project lead by Gupta N. at the MeerKAT radio telescope, for a blind search of HI and OH absorbers at $z < 1$, that is already giving some promising results regarding the fraction of DLAs associated to the CNM phase (Gupta et al., 2021a). They found a HI 21-cm absorption detection rate of 30%,

which implies a warm neutral fraction associated to DLAs even higher than what found by Kanekar et al. (2014): from the measurement of DLA spin temperature they found that only two out of 23 DLAs at $z > 1.7$ have T_s values indicating CNM fraction $> 20\%$.

If we assume that most DLAs ($\sim 80\%$) are associated to the warm neutral phase, then the environment of DLAs in galaxies does no more coincide with the birthplace of molecular hydrogen and this could explain the observed lower covering factor of strong H2 absorbers.

These findings highlight the necessity of including a detailed treatment of the warm neutral phase in our model for the galaxy formation and evolution. This would obviously impact different aspects of the modeled galaxy evolution, because it will change the depletion time of the gas and therefore the star formation rate at different redshift.

4.6 Testing the importance of the ejected gas for DLA statistics

We have shown in our previous work that the ejected phase contains enough gas to explain the observed amount of HI gas in the Universe, if we assume that a part of it gives origin to the observed dense Lyman-α absorbers (in particular to WNM associated with HI absorbers). Starting from those results we have decided to test whether the ejected gas associated to galaxies could give a not negligible contribution to the DLA covering fraction. If this is the case it could potentially solve the discrepancy between the predicted and observed CDDF of $H2$ absorbers. Before assuming a model to estimate the warm and cold neutral hydrogen fraction in the ejected component, or outflow component, I have considered all the ejected gas as HI. And then I have studied if this ejected HI associated to each model galaxy, contributes significantly to the DLA covering fraction. To estimate the HI column density associated to the ejected component I have to model the spatial distribution of the ejected gas and I have assumed the geometry of a spherical shell, for simplicity (this geometry is often adopted in other works on DLAs: Fynbo et al., 2011; Krogager & Noterdaeme, 2020). However, it's important to highlight that this assumption is not completely physically motivated and outflows are often observed to follow a biconical distribution (Veilleux et al., 2001; Veilleux et al., 2020).

The spherical shell is completely defined by two radii: the internal radius (R_{in}) and the external radius (R_{out}).

The first thing to test is whether the contribution of the ejected gas to the DLA covering fraction depends on the radial extent of the spherical shell hosting the ejected gas. I have assumed three different combinations of values for (R_{in}, R_{out}) :

1. (R_{in}, R_{out}) $= (0.2, 1.0)R_{200}$
2. (R_{in}, R_{out}) $= (0.5, 1.5)R_{200}$
3. (R_{in}, R_{out}) $= (0.2, 1.5)R_{200}$.

In fig. 4.27 I show the predicted numbers of DLAs in bins of column density for different runs of the **GAEA-RPS-mod-BR-evo**P_0model in which I change the radial extent of the shell hosting the ejected gas. We can observe

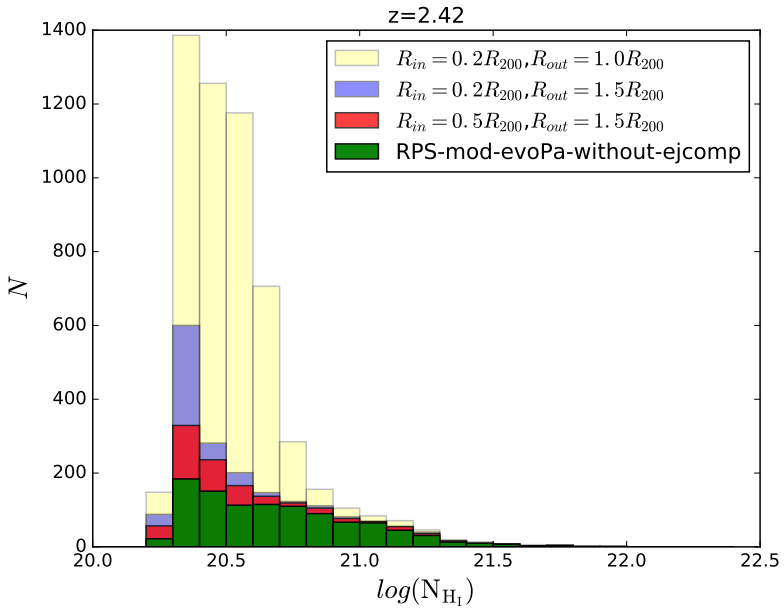

Fig. 4.27: Histogram of DLA HI column densities for the **GAEA-RPS-mod-BR-evo**P_0 run over the MSII box, with different spatial distributions of the ejected component.

that the run that predicts a greater contribution of the ejected component to DLAs is the one with the smallest spherical shell (yellow), despite the decrease in the number of shells intersected by l.o.s. This happens for 2 reasons:

1. decreasing the volume of the spherical shell increases the surface density of HI (inversely proportional to $[(R_{in} + R_{out})^2 - R_{in} * R_{out}]$) encountered by the l.o.s,

2. when the value of R_{in} is smaller, the mass of ejected HI in the shell more often overlaps the HI housed in the internal regions of the discs.

Then, I tested only for the **GAEA-RPS-mod-BR-evo-**P_0 run with the smallest ejected spherical shell, the effect of not considering all the ejected gas as neutral atomic hydrogen but only the gas which satisfies the temperature condition $T < 10^5$K. Since we are not able to assign self-consistently a temperature to the gas in the ejected phase in the GAEA model, I have applyed the fitting functions derived from the observed properties of simulated outflows in the FIRE-2 simulation (Pandya et al., 2021) to estimate the cold ($T < 10^3$K) and warm ($10^3 < T < 10^5$K) mass fraction associated to the

ejected gas component. According to these fitting functions, the cold gas fraction depends only on the redshift, as shown in Eq. :

$$f_{cold} = 10^{-3.1}(1+z)^{3.1} \tag{4.22}$$

while the warm fraction depends on the stellar mass of the galaxy ejecting the gas:

$$f_{warm} = \begin{cases} 10^{-0.5}(M_\star/10^{10.5}M_\odot)^{-0.09}, & \text{for } M_\star \leq 10^{10.5}M_\odot \\ 10^{-0.5}(M_\star/10^{10.5}M_\odot)^{-2.0}, & \text{for } M_\star > 10^{10.5}M_\odot \end{cases}$$

Before going to derive the DLA observables for this model we can test whether the predicted cosmic HI density associated to galaxies is sufficiently high to be in agreement with the estimates of HI comoving density derived from DLA observations, if we consider the HI settled in the ISM plus the HI in the warm/cold fraction of the ejected gas component, associated to all the simulated galaxies.

In Fig. 4.28, I show the predicted HI density associated to the ISM in galaxies ($\Omega_{HI,ism}$), the predicted total HI density, given by the HI in the ISM plus the ejected component associated to galaxies ($\Omega_{HI,tot}$, where $M_{HI,ejected} = f_{warm} \cdot M_{ejected} + f_{cold} \cdot M_{ejected}$) and the cold gas plus ejected gas density ($\Omega_{CG+ejected}$), with solid, dashed, and dot-dashed lines, respectively, for the **GAEA-RPS-mod-BR-evo**P_0 run over the MSII box.
As shown in figure, the comoving total HI density, associated to the ISM and the ejected gas in galaxies, is always higher than the HI cosmic density estimates coming from DLA observations and it flattens at $z > 1.5$, following a slope more similar with the one of data, while the cold gas plus ejected gas density decreases slowly at $z > 2$. Therefore we could expect that this run will reproduce better the evolution of $\Omega^{\rm HI}_{\rm DLA}$ with redshift and I move to the estimation of DLA statitics.

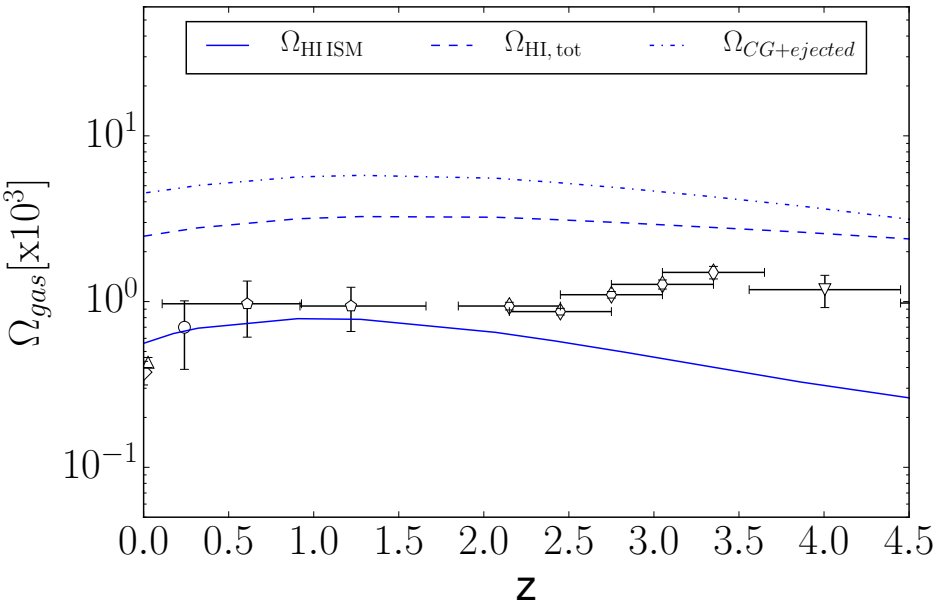

Fig. 4.28: The comoving density of $HI_{\rm ISM}$, $HI_{\rm ISM+ejected}$, and cold gas+ ejected gas associated to galaxies in the MSII box, for the **GAEA-RPS-mod-BR-evo** P_0 run.

The first DLA observable which I investigated is the DLA CDDF. In Fig 4.29 I show the average CDDF in the redshift range $2 < z < 3$, predicted by the **GAEA-RPS-mod-BR-evo**P_0 model, with (blue) and without (red) the HI in the ejected component. In the top panel I show the predictions associated to the run where I take as HI all the gas in the ejected phase while in the bottom panel I consider as HI "ejected" only the warm and cold neutral gas associated to the ejected component. Both the two runs overpredict the number of DLA absorbers contributing to the low-N_{HI} end of the CDDF, but in the second case considered the overprediction impacts only at $log(N_{HI}) < 20.5$, where we have more systematic errors which could bias observations, even if we must notice that the statistics is very large, and it should reduce the effect of systematic errors. It is worth noticing that considering also the ejected HI does not affect very much the high-N_{HI} end of the CDDF. The haloes which contribute mostly to the high-N_{HI} end of the CDDF are the more mass haloes, therefore we expect that the ejected component in this halos do not give an important contribution to the DLA covering fraction.

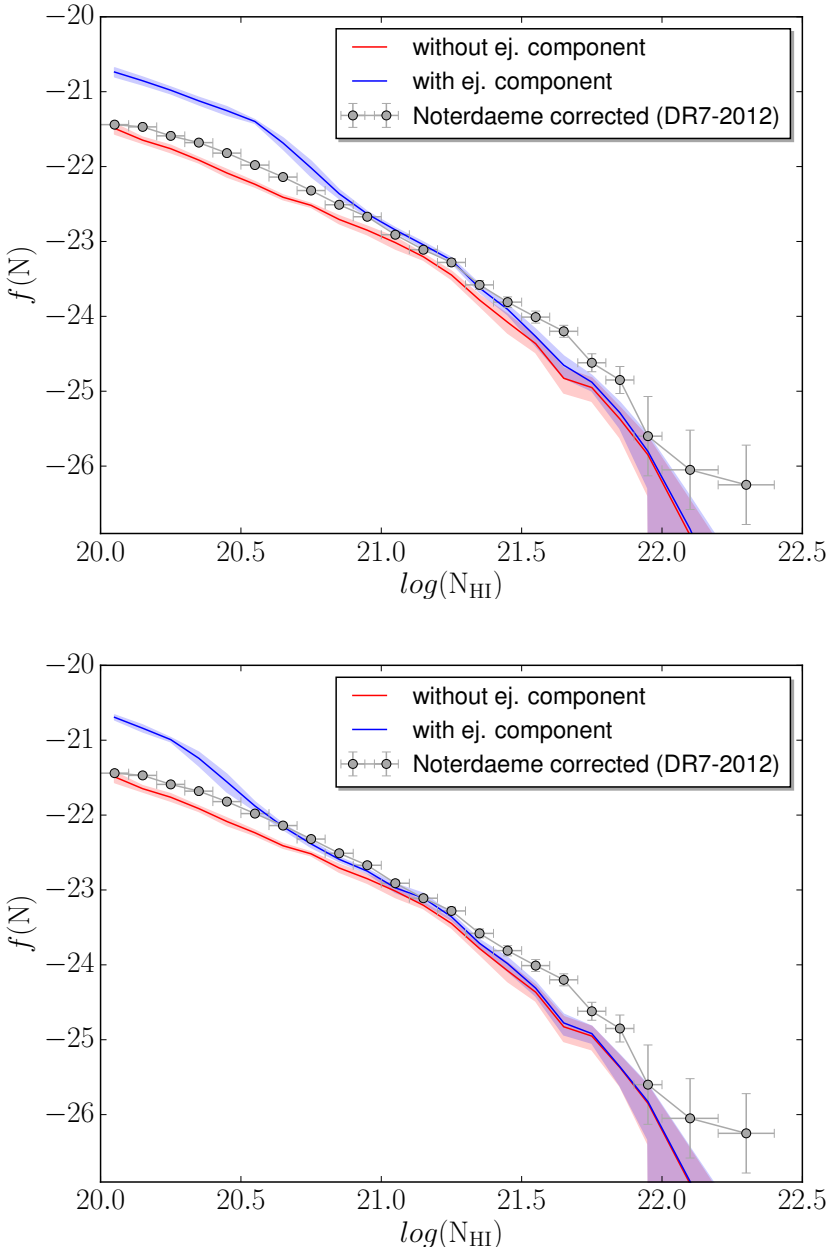

Fig. 4.29: Average predicted CDDF for $2 < z < 3$ from the **RPS-mod-BR evo** P_0 model. In the top panel I show the results from the run where I consider all the gas in the ejected component as HI, while the bottom panel refers to the run where $M_{HI,ejected} = f_{warm} * M_{ejected} + f_{cold} * M_{ejected}$. Data points refer to the observed total CDDF in the redshift range $2 < z < 3$ (Noterdaeme et al., 2012)

The reason for this can be understood looking at the dependence of the ejected mass on the halo mass in our model, shown in Fig. 4.30. The average amount of the ejected gas decreases going towards larger halo masses and this is due to the shorter re-accrection time at larger halo masses.

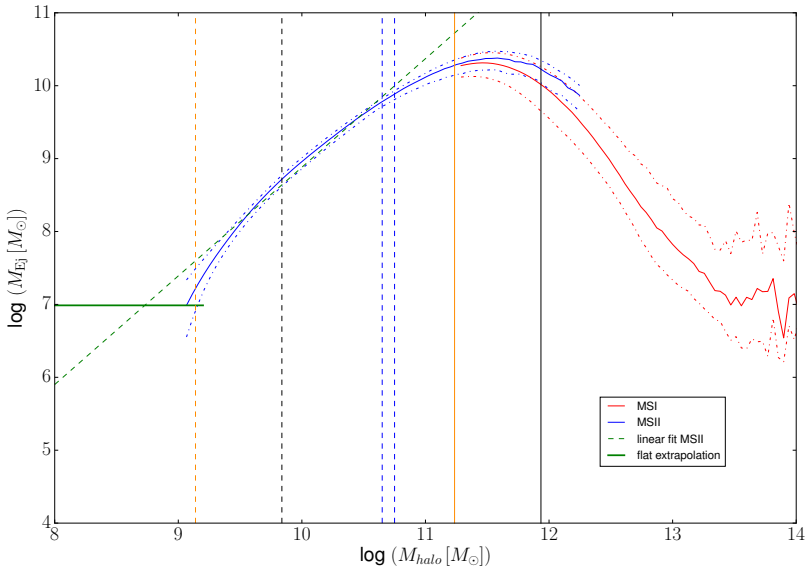

Fig. 4.30: The median ejected mass of central galaxies binned in halo mass, at $z = 2.07$, for the **GAEA-RPS-mod-BR-evo** P_0 run on the MSII box. The vertical solid and dashed lines correspond to 150 times and 1000 times, respectively, the DM particle mass for the MSI and MSII simulations.

The second tested observable is Ω_{DLA}, and its evolution with redshift. In Fig 4.29. I show the comparison between the predicted Ω_{DLA} by the **GAEA-RPS-mod-BR-evo** P_0 run, with and without the HI in the ejected component. In the top panel I show the results associated to the run where I take as HI all the gas in the ejected phase while in the bottom panel I consider as HI "ejected" only the warm and cold neutral gas associated to the ejected component.

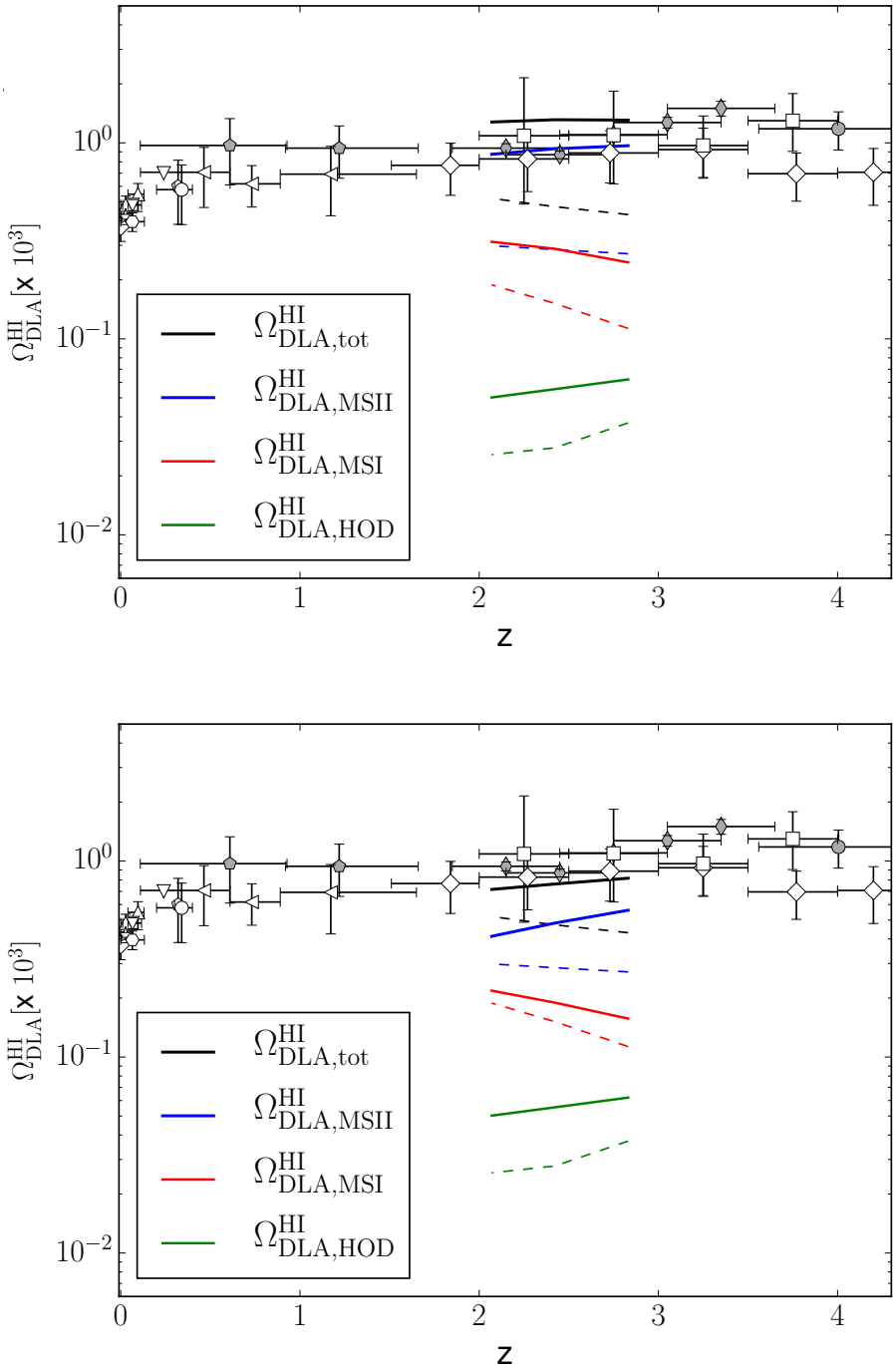

Fig. 4.31: Comparison of the predicted Ω_{DLA} from the **RPS-mod-BR evo** P_0 model, with and without the ejected HI. Similarly to the previous plot the top panel shows the results from the run where I consider all the gas in the ejected component as HI, while the bottom panel refers to the run where $M_{HI,ejected} = f_{warm} * M_{ejected} + f_{cold} * M_{ejected}$. Filled grey data points refer to the observed Ω_{DLA} estimates collected by Crighton et al. (2015), while the white data points are taken from the recent review by Péroux & Howk (2020).

We can observe that, if we consider also the contribution of the HI ejected to the HI comoving density associated to DLAs, the total comoving density increases at all redshifts of interest, and in particular at higher redshift, suggesting a change in the shape of the Ω_{DLA} evolution with z [7] . The observed change in the Ω_{DLA} evolution with z is a very important result, since we have shown in the past that our model could not reproduce the evolution of Ω_{DLA} at $z > 2$ with any combinations of changes in cold gas masse, cold gas scale radii, and cold gas partitioning schemes. The past model versions considered predict always an Ω_{DLA} decreasing with z, while the observation show a slow increase of Ω_{DLA} towards higher redshifts.

Besides, comparing the prediction in the top panel with the one in the bottom panel we can observe that the slope of the Ω_{DLA} evolution function is steeper when we apply the FIRE fitting function to derive the HI mass associated to the ejected component.

From the comparison of the predicted evolution for Ω_{DLA} with the predicted evolution for $\Omega_{\mathrm{HI},tot}$ (shown in Fig. 4.28), we could deduce that the adopted fitting functions, to estimate the warm/cold gas fraction associated to the ejected component, contribute only in part to the change in slope of the Ω_{DLA} evolution function, while another mechanism which affects equally the evolution with z of Ω_{DLA}, and it could be the effect of the geometrical distribution of the ejected HI, and its dependence on R_{200}. However we need further investigation to corroborate this hypothesis. Besides, I plan to test the effect of assuming other geometrical distribution for the ejected HI, on the predicted DLA statistics.

4.7 Conclusions

In conclusion, I have shown in this Chapter that moving from an early Reionization scenario to a late one does not lead to tremendous differences in terms of the predicted galaxy properties at $z < 4$, in the GAEA framework. Besides, after testing different prescriptions for the UVB feedback the one which better reproduces more up-to-date results from hydro-dynamical simulations (the Maccio's prescription) increases only partially the agreement between model predictions and DLA observations: we recover a better shape

[7] However, we need to extend the redshift range investigated, towards higher z, to recover the shape of the predicted evolution for Ω_{DLA}

of the DLA CDDF but we are not able to diminish sensibly the offset with this new prescription.

The last tested modification to the model prescriptions regards the partitioning of the cold gas in the ISM. Motivated by the results of the theoretical work by Feldmann et al. (2012) and of the observational review by Péroux & Howk (2020), I have tested a modified BR prescription for the cold gas partitioning in which I consider a value for P_a evolving linearly with redshift (as described in eq. 4.6). With this new prescription, we are able to recover the observed DLA CDDF in the redshift range $2 < z < 3$ in the framework of the GAEA-RPS model, and also the distribution of DLA metallicities whether we take into account also the metals in the outflows (as shown in Fig. 4.20.

In the last section of this chapter I have also investigated the predictions of our new reference GAEA-RPS run for modeled galaxies and DLA properties. In particular I have focused on the molecular content of modeled galaxies and DLAs. And I have discovered that, for the ASPECS adopted value of α_{CO} this model underpredicts the molecular hydrogen content associated to galaxies at all redshifts higher than 0. For $\alpha_{CO} = 2.0 M_{\odot}[K\, kms^{-1} pc^{-2}]^{-1}$the model begins to underpredict the cosmic H_2 density at $z > 1.5$, just before the Peak Star Formation Epoch (as shown in Fig. 4.24). On the contrary, if we assume $\alpha_{CO} = 0.8 M_{\odot}[K\, kms^{-1} pc^{-2}]^{-1}$the model predictions agree with observation at $z > 1$ but it predicts too large estimates of $\rho(H_2)$ in the Local Universe, even if it is worth noticing that the low-redshift point of De Carli et al. 2019 is deemed unreliable (and it is important highlight that the last adopted value for α_{CO} does not apply likely to the entire ASPECs sample, for the reasons discussed in Popping et al. (2019b)). Moreover the shape of the H_2 density evolution with z is not into agreement with the observed one, because we do not observe a peak in $\rho(H_2)$ around $z \sim 2$ for any of the assumed conversion factors from L_{CO} to $M(H_2)$, and instead the maximum value of $\rho(H_2)$ is always around $z = 0.5$, where we observe a flattening in the evolution of $\rho(H_2)$. The observed evolution of $\rho(H_2)$ is in part influenced by our choice of the prescription for the cold gas partitioning, which assigns more atomic hydrogen (and consequently less H_2) to galaxies at higher redshift with respect to the classical BR prescription, but also other runs of the GAEA-RPS model and the past model version, GAEA-X17, underpredict the cosmic molecular density at higher z, and in particular at $z \sim 2$ (as shown in Fig. 4.12 of this Chapter and in Fig. of Xie et al. 2020), and this could be

the combined effect of several ingredients which regulate galaxy evolution in our model (e.g star formation efficiency, stellar feedback mechanism). The fact that at high redshift, for the more reasonable values of α_{CO}, this model under-predicts both the cosmic densities of the molecular hydrogen and the atomic hydrogen, constrained by different types of observations (CO and DLA observations, respectively) may suggest that a possible cause of this underprediction is an excessive star formation activity but the model seems to predict reasonable values of star formation rates at all redshifts (as shown in Hirschmann et al., 2016). Therefore we need to investigate further other possible causes for this finding.

It is important to highlight that the same GAEA model overpredicts the number of dense molecular absorbers associated to DLAs at $2 < z < 3$ (as shown in Fig. 4.26). This could seem in contrast with the results regarding the cosmic H_2 density associated to galaxies, but it's worth noticing that the molecular hydrogen associated to DLAs trace only a little part of the total molecular hydrogen amount in the Universe. Besides, the GAEA model discrepancy with the observations by Balashev & Noterdaeme (2018) is likely influenced by an over-simplification in the modeling of DLA systems. In particular we do not model in GAEA the warm gas, neither in the ISM nor in other galactic components. Observations suggest, instead, that warm neutral gas, originated from ISM/CGM or outflows, contributes in a not negligible way to the DLA covering fraction. Besides, considering the warm/cold HI associated 'a posteriori' to the ejected gas phase, even if with a very simple model, changes significantly the DLA CDDF in the redshift range of interest, as shown in the last section of this Chapter. Therefore, I expect that modeling in a self-consistent way the warm HI in the ISM/CGM (in particular in the ejected component) we can improve the DLA statistics predicted by the GAEA-RPS model, coming also into a closer agreement with the observations on H_2 dense absorbers.

In the near future I intend to further investigate the relation between different geometrical distributions of the ejected HI and the predicted DLA statistics. Besides, possible implementations for self-consistently modeling the warm neutral gas in GAEA are under investigation and will be discussed in the next Chapter.

Chapter 5

Discussion of the thesis results and future developments

5.1 Discussion of thesis results

In this thesis work, I have investigated the connection between the densest absorbers (DLAs) in the Intergalactic Medium (IGM) and galaxies at different redshifts, taking advantage of state-of-the-art semi-analytic model of galaxy formation GAEA, developed at the Observatory of Trieste by De Lucia's research group (De Lucia et al., 2014; Hirschmann et al., 2016; Xie et al., 2017). This is not the first study on DLAs based on semi-analytical models (Berry et al., 2014, 2016). However, differently from the previous studies, our model offers the possibility to populate the simulated box with haloes extracted from the merger trees of a cosmological DM only simulation (the Millennium simulation). This provides the connection between the simulated DLA properties, the baryonic history of the DLA host galaxy and the host halo merger history. Moreover, the GAEA model accounts for a non-instantaneous recycling of gas, metals and energy.

The understanding of the connection between DLAs and galaxies has two important benefits. On the one side it helps to better characterize the physical properties of the IGM. On the other side, it can shed light on the processes which regulate the baryon cycle in galaxies. DLAs are expected to typically arise from the neutral gas settled within $30 - 50$ kpc from their host galaxies (Rahmati et al., 2015; Rubin et al., 2015). It is still not clear if most

of them originate in the galactic ISM or whether outflows/inflows contribute equally to their covering fractions (van de Voort et al., 2012; Pehlivan Rhodin et al., 2019), and how the relative contribution of ISM and CGM to σ_{DLA} evolves with z. It is also expected that the filament contribution to DLA cross-section would become important at high redshfit but it should happen at $z > 3$ (Rahmati et al., 2013).

In my first work, presented in Chapter 3, I have tested the hypothesis of an unique association of DLAs systems and neutral hydrogen in the ISM of galaxies, for absorbers origated at $z < 3$. As a first step, I have compared the predictions of the GAEA model, in particular of the GAEA-X17 (BR) run, with recent measurements of two of the most tested DLA observables: the DLA column density distribution function (CDDF) and the comoving HI density. I restricted my analysis to the redshift range $2 < z < 3$, where observational data are more accurate, and the galaxy star formation activity is at its maximum. As a result of the comparison, we discovered that our fiducial model predicts the correct shape of the column density distribution function, while its normalization falls short (about a factor 2.5) of the observations, with the discrepancy increasing at higher redshift. Modifying 'a posteriori' the radius of cold gaseous disks and the cold gas mass associated to simulated galaxies, by multiplying these quantities by a factor 2, it is possible to recover almost all the missing HI, up to $z \sim 2.5$. At higher z, the HI cosmic density predicted by our model declines while the observed $\Omega^{\rm HI}_{\rm DLA}$ stays almost flat. This could be seen as a hint for missing neutral gas associated to other galactic/extragalactic components (e.g. outflows, filaments).

In the same work, I have shown that the GAEA-X17 model predicts a significant excess of low-metallicity DLA systems with respect to observations (De Cia et al., 2018). The average cosmic DLA metallicity (Ω_Z), weighted over the column density, follows, instead, the same redshift evolution of the observational measurements but it is slightly higher than observed values. The predicted Ω_Z becomes compatible with observations once we account for a modest radial metallicity gradient, in particular with the universal slope derived by Christensen et al. (2014a).

I suggested possible ways to improve the agreement with observations of DLA statistics at intermediate redshift without destroying the agreement with other observational constraints:

1. modifying the assumed reionization scenario (bringing it more in agree-

ment with recent Planck results) and the adopted prescription for the Reionization feedback on galaxies (or UVB feedback).

2. modifying the treatment of the cold gas angular momentum, which could lead potentially to biased small radii of the galactic gaseous disks, and could impact the simulated DLA cross-section.

3. modifying the assumed cold gas partitioning scheme,in GAEA-X17 (BR). Since the BR law has been derived from observations in the local Universe, it could loose validity at high redshift, affecting the HI content of all the modeled galaxies in the reshift range of interest.

In my second work, described in Chapter 4, I have investigated all these hypotheses, adopting a new version of the GAEA model, the GAEA-RPS model (Xie et al., 2020b), which contains different improvements w.r.t. GAEA-X17. The model adopts an updated treatment of the cold gas angular momentum in galaxies, and it considers both the hot gas and cold gas stripping processes happening in dense environments. As shown in Xie et al. (2020b), this model predicts an average higher HI mass associated to central galaxies, w.r.t. to GAEA-X17. This can increase the agreement between the GAEA model predictions and DLA observations, since we have shown that in our model DLAs originate mostly from central galaxies. The first step to investigate the DLA-galaxy connection is to study the cold gas/HI content of galaxies, its evolution with redshift, and the relation between the cold gas content and other galaxy properties. With this purpose in mind, I have compared the HI content of model galaxies to observational data, analyzing several statistics: the HI/H2 mass function, the HI-stellar mass relation, and the HI-halo mass relations. I have tested how the predictions regarding these observables change under different assumptions.

As a starting point, we assumed a late reionization (LR) scenario, in agreement with more recent observational constraints on Reionization, and in particular with the ones coming from the Planck mission (Planck Collaboration et al., 2018). In the old reference GAEA model (X17, Xie et al., 2017) was instead assumed an early reionization (ER) scenario. The first section of Chapter 4 is devoted to this study. Moving from a ER scenario to a LR one induces a small change in the stellar mass-halo mass and HI mass-halo mass relation at low masses ($M_{200} < 10^{10.5} M_{\odot}$), and it slightly reduces the HI cosmic density in our simulated Universe, for $z < 4$. The latter result is a

consequence of the fact that the adopted prescription for the UVB feedback (Gnedin, 2000) predicts a bigger value for the filtering mass at low redshift if Reionization ends later.

I also tested the effects of three different prescriptions for the UVB feedback on the atomic and stellar content of the galaxies modeled by the GAEA-RPS model, in particular for the BR run (Blitz & Rosolowsky, 2006). As shown in Fig. 4.9 and 4.10 the Kim's prescription (Kim et al., 2015) leads to a SMF and HIMF incompatible with observational constraints from the local Universe. The GAEA-RPS runs based on the prescriptions for UVB feedback by Macciò et al. (2010) or Kravtsov et al. (2004) give instead realistic SMF and HIMF at $z = 0$. With the Macciò prescription, theoretically motivated by the results of the work by Okamoto et al. (2008b), the **GAEA-RPS-BR** model also predicts a better shape of the simulated DLA CDDF, for $2 < z < 3$ (as shown in Fig. 4.11). However, this modification does not solve the normalization discrepancy between the simulated CDDF and $\Omega^{\rm HI}_{\rm DLA}$ with the observed ones. Then, I tested the effect of different cold gas partitioning schemes, in particular the GK (Gnedin & Kravtsov, 2011), KMT (Krumholz, 2012), BR (Blitz & Rosolowsky, 2006) and modified BR scheme (described in details in Section 4.4) on the properties of model galaxies. The results of this investigation are reported at the end of Section 4.4. If we assume a modified BR prescription, where the normalization of the BR law decreases linearly with z (see Eq. 4.21), the cosmic HI density increases going from $z = 0$ to $z = 1$ and the predicted DLA CDDF, in the redshift range, is on average in agreement with the observed CDDF, and the predicted SMF and HIMF satisfy the current observational constraints, while the GAEA-RPS-mod, when adopting the KMT and GK schemes for the CG partitioning in the ISM, predicts a SMF in disagreement with observations in the high-mass end, and a HIMF with a bump at intermediate HI masses. Therefore, we adopt the **GAEA-RPS-mod-BR-evo-$\mathbf{P_0}$** as our new reference model and test the prediction of this model for other DLA and galaxy properties (e.g. metallicity and H_2 content). The first test regards the metallicity distribution of DLAs. To estimate the amount of metals associated to DLAs in our model, I have adopted two different procedures: i) associate to the DLAs only the metal content predicted by our model for the cold gas phase, and ii) consider also the metals in the hot gaseous phase and in the ejected phase. After this step, in order to have an estimate of the abundance ratio $[Fe/H]$ for each simulated DLAs, I distributed spatially the iron content

following the same exponential profile of the cold gas. I have considered also a radial metallicity gradient with the universal slope measured by Christensen et al. (2014b). The new reference model predicts a 2D distribution for the [Fe/H] VS $N_{\rm HI}$ which differs significantly from the observed one, if we consider only the metals associated to the cold gas phase (see Figg. 4.19 and 4.20). When considering the metal content of CG+HG+Ej gas, the predicted 2D distribution comes in agreement with the observed one, even if our model still predicts systems with a metallicity under the observed 'metallicity floor'. In section 4.5, I discuss another test of the new reference model regarding the molecular hydrogen associated to model galaxies and simulated DLAs. The most used conversion factor from CO luminosity to H_2 mass (α_{CO}) is based on observations in the local Universe and impacts a lot the derived estimates for the H_2 cosmic density and the H2MF. I followed the approach described in the work by Popping et al. (2019b), considering multiple values for α_{CO}. I showed that the predictions from our new reference model are in agreement with observations at high z assuming $\alpha_{\rm CO} = 0.8 M_{\odot}/(K$ km/s $pc^2)$, while the model overpredicts the amount of cosmic H_2 for z~ 0. It's worth noticing that it is quite unlikely that this value of the conversion factor describes the relation between CO luminosity and molecular hydrogen amount in all the galaxies observed with ALMA (Decarli et al., 2019). Besides I have investigated the predictions of this model for the molecular hydrogen associated to DLAs. I have shown that the new reference model predicts a CDDF(H_2) almost one order of magnitude offset high w.r.t. the one derived by Balashev & Noterdaeme (2018). This could be due to different aspects neglected in our model: the clumpiness of the molecular hydrogen in the ISM/CGM, the multi-phase structure of the ISM/CGM, and in particular the existence of a warm phase. This latter could potentially host the majority of the observed DLAs, according to some observational studies (Kanekar et al., 2014; Gupta et al., 2021b) . Moreover, I have tested the contribution to the DLA cross-section coming from neutral gas in the ejected gas component associated to simulated galaxies in the **GAEA-RPS-mod-BR-evo**P_0 model. I have shown that this component contributes predominantly to the low NHI-end of the CDDF and that the geometry assumed for the distribution of the 'ejected' HI is important. I have tested different volumes for the ejected component, assuming a spherical geometry, for simplicity. Nevertheless, I expect that assuming a bi-conical geometry or a sheet-like geometry will impact in a

not negligible way the model predictions, and I plan to test these hypotheses following a more physically motivated approach, as described later on.

In conclusion, I have found that, at least in the framework of semi-analytical models, a change in the Reionization timeline, moving from a ER scenario to a LR, does not impact significantly the model predictions on the HI cosmic density while important differences in the characteristic masses at which haloes retain only half of the universal baryon fraction, as a consequence of the UVB feedback, are associated to significant changes in the properties of model galaxies. Besides, differently from what it has been shown in Xie et al. (2017) for the GAEA-X17 model, in the GAEA-RPS model, assuming different schemes for the cold gas partitioning brings to different predictions for the SMF and HIMF, at low and high redshift, and the only model, among different RPS runs with modifed reionization history and UVB feedback prescription, which reproduces the observational constraints on galaxies and DLA properties, in the redshift range $0 < z < 2.5$, is the **GAEA-RPS-mod-BR evo-$\mathbf{P_0}$** model, our new reference model.

5.2 Future plans

The study of DLA connection with galaxies can potentially shed light on the physical conditions of the multi-phase turbulent ISM and CGM of high-z galaxies (Krogager & Noterdaeme, 2020) but some fundamental questions regarding the origin of these systems need to be solved (as highlight by van de Voort et al., 2012; Bird et al., 2015; Pehlivan Rhodin et al., 2019; Balashev & Noterdaeme, 2018), in order to advance our understanding of the physical processes ruling the gas cycle and star formation in galaxies: what is the covering fraction of DLAs in outflows/inflows? What is the origin of DLA kinematics and metallicity distribution? What fraction of the total neutral hydrogen content of galaxies is in the warm neutral phase at different z? With my future studies, I aim to answer these questions and help in developing a more complete picture of the cold gas cycle in galaxies, for $z < 5$.
This is the ideal time for investigating the DLA/galaxy connection and, more in general, the cold gas cycle in galaxies, since present-day telescopes are collecting observations that put constraints on the cold gas distribution in/around galaxies: wide-field quasar surveys in optical/NIR band, observations of DLA environments (e.g. MUSE at the VLT), with a particular focus

on the relation between the DLA cross-section and intra-group gas or neutral inflows/outflows in galaxies, ALMA observations of emission lines in the quasar proximity, which allow to probe the energetics, geometry and chemistry of quasar outflows, and 21cm absorption and emission line observations, that constrain cold gas properties of galaxies both in the local and in the more distant Universe.
In particular, Lyman alpha emission observations of DLA fields allow us to probe the connection between HI gas and star formation in galaxies at intermediate/high redshift, while the comparison of DLA statistics with the statistics of dense 21cm absorbers, collected with advanced radio telescopes and arrays (in particular, MEERKAT and ASKAP, both SKAO precursors), will shed light on the warm neutral fraction in DLAs and its evolution with redshift (e.g. Kanekar et al., 2014; Gupta et al., 2021b).

To get a coherent picture of the cold gas cycle in simulated galaxies I will need to trace simultaneously the ISM, CGM, and inflows/outflows of molecular and atomic gas, disentangling the resolution effects on the gas properties from the dependence of gas properties on different prescriptions adopted to describe the cold gas accretion, the cold gas partitioning, the star formation, the stellar/AGN feedback and the metal treatment in simulated galaxies.

High-resolution hydro-dynamical simulations are essential tools to study the cold gas cycle because they can trace the kinematics and spatial distribution of cold gas in different galactic environments (e.g. ISM, CGM, and inflows/outflows) and describe in a self-consistent way the multiphase structure of ISM/CGM (Hu, 2000; Richings & Schaye, 2016; Hill et al., 2018). However, they have not been successful yet in disentangling the effects of the assumed sub-grid physics on the cold gas properties, in particular at intermediate column density, from resolution effects and numerical artifacts (e.g. Kauffmann et al. 2019). Moreover, their high computational cost does not allow to probe the cosmological volume needed to sample properly the halo cosmic variance, which impacts significantly the DLA statistics (as shown in my first work). A possibility to overcome the latter problem is to use the GAEA semi-analytical model, with prescriptions for neutral gas in outflows/CGM tuned on the TNG50 simulation results, coupled to a cosmological N-body simulation like the MSI simulation. I will focus on the halo mass range $M_{200} \in (10^9 - 10^{12}) M_{\odot}$, which is well resolved in TNG50 and covers

all the halo masses contributing in a not negligible way to the DLA CDDF (as shown in my first work: Di Gioia et al., 2020). I will follow the baryonic evolution of ~ 10 galaxies for each halo mass bin, sampled from different environments, classifying them in terms of AGN and stellar feedback properties. Then, I will derive a 3D profile for the distribution of cold/warm gas in the outflow/inflow component associated to these galaxies, choosing the fitted distribution with the best chi-square among the ones with a compatible number of parameters. In particular I plan to test whether the cold and warm gas particles around galaxies, and within $R = 1.5R_{vir}$, follow a spherical or bi-conical distribution, and if their spatial distribution changes moving from galaxies where the stellar feedback is dominant to galaxies where the AGN feedback is dominant.
Once I have derived the best-fit 3D distribution for cold gas associated to outflows and inflows in galaxies (dominated and not dominated by AGN feedback) I will adopt this 3D distribution for the cold gas associated to the ejected component in GAEA and re-estimate the predicted DLA statistics at different redshifts. This study would be really important to characterize the relation between feedback processes in galaxies and the spatial distribution of cold gas in/around galaxies, and how this impacts on DLA statistics.
I expect that the outflow contribution to DLA cross-section will dominate the inflow contribution because otherwise it would be difficult to reproduce the observed distribution of metallicities associated to DLAs. If outflows contribute to most of the cold gas outside the galactic ISM, we could expect a distribution more similar to bi-conical one. There are already some observations of cold gas in outflows that suggest that the bi-conical distribution as more likely (Bouché et al., 2012; Kacprzak et al., 2012, 2015; Zabl et al., 2020; Veilleux et al., 2020), but there is not yet a sufficiently large statistics.

However using an hydro-dynamical simulation with a box side length of 50 cMpc, like TNG50, does not allow to reach the resolution necessary to probe all the important baryonic physical processes which regulate the content and distribution of cold gas in the CGM (as shown in van de Voort et al., 2019). To advance this study, reaching the required resolution with a reasonable computational time/cost, I plan to build a suite of zoom-in high resolution hydro-dynamical simulations for ~ 30 halos in the mass range $(10^9 - 10^{12})M_{\odot}$, settled in different Lagrangian regions of a cosmological simulation, selected to probe in a statistical way the cosmological volume

(sampling regions with different large-scale densities, from voids to cluster regions). The zoom-in method consists in improving the resolution of specific sub-volumes of a lower-resolution cosmological simulation. The number of particles in these sub-volumes is increased to reach the desired resolution, (a likely desirable high-resolution simulation for DLA studies would adopt a resolution mass of the order of $10^4 M_\odot$ for the DM particle and 10^3 for the gas particle) and outside the selected regions the resolution is progressively degraded, saving computational time, while preserving a correct description of the large-scale tidal field. For this type of study it would be suitable a code similar to the one at the basis of the more recent simulation developed by the group lead by Dr F. Van De Voort (van de Voort et al., 2019). Of course, this type of studies is really expensive in terms of time ($> 50k$ CPU hours on a machine like MARCONI at CINECA) and computational costs at the moment but it would be feasible once the European Exascale HPC Research computing facility [1] will be in place.

Another path that it is worth pursuing, in parallel to the previous one, is coupling GAEA with the merger trees extracted from an high-resolution cosmological DM only simulation and introducing more complex prescriptions for the multiphase ISM and CGM. These prescriptions should consider the clumpy nature of the molecular hydrogen (as done in Popping et al. 2020) and also the presence of a warm neutral phase in the ISM, associated to a temperature range $T_{WNM} \in (10^{3.5}, 10^5)K$, and observed to represent the 60% of the neutral gas in the diffuse phase of the ISM in the MW (Kalberla & Kerp, 2009; Gerin et al., 2015; Martin et al., 2015). At the moment in the GAEA model we consider only the neutral gas in the ISM, and in particular only the cold neutral phase of the ISM, partitioned into atomic and molecular hydrogen, plus He, and the star formation depends only on the molecular hydrogen density, which is distributed continuously on galactic disk. The ISM modeling can be refined taking into account the turbulent nature of the ISM, very important at high redshift, and the intrinsic clumpiness of molecular hydrogen settled in dense ISM regions. To account for the effects of turbulence in high-redshift galaxies I plan to introduce in the GAEA model a parametrization for distinguish warm and cold neutral gas in the ISM, depending on the value of the interstellar radiation field and on the gas specific angular momentum (for example using the model developed by

[1] https://www.etp4hpc.eu/euexascale.html

Bialy & Sternberg (2019)). The introduction of a two-phase neutral medium could impact in a non-trivial way the star formation histories of the modeled galaxies since it will change the depletion time of the neutral gas, therefore the treatment for a 2-phases neutral medium in the ISM, should be calibrated in order to keep the agreement with observations on the predicted galaxy properties, like stellar mass,SFR, cold gass mass, radius of the galactic disk, amount of metals.

In parallel I plan to model the clumpiness of the molecular hydrogen, adopting an approach similar to the one described in Popping et al. (2019a). As starting point I will divide each model galaxies in 30 annulii where I will estimate the ionized and neutral fraction of gas. Then, from the neutral gas fraction I will derive the warm and cold neutral fraction, in each annulus. After that I will divide the cold neutral gas into atomic and molecular (H+ He). After having estimated the molecular fraction of the cold neutral gas in each annulus, I can subdivide the total molecular mass among the molecular clouds, which will follow a power-law mass spectrum. Each molecular cloud would be defined by its molecular mass, size, internal density profile and chemistry. The sizes of molecular clouds (defined uniquely by their radii) are derived by applying the virial theorem, where we consider the external pressure acting on the galactic disc at the position of the molecular cloud, and the gravitational force. To model the internal density profile I will adopt the Plummer density profile (firstly suggested by Whitworth & Ward-Thompson, 2001), as Popping et al. (2019a) has shown that adopting this profile to describe the internal mass distribution in molecular clouds , in the SC model, leads to a better agreement between the predicited CO luminosities and the observed one. The chemistry of molecular clouds is regulated by the heating processes driven by the local interstellar radiation field ($G_{\rm UV}$) and by cosmic rays (CR). To model the intensity $G_{\rm UV}$ field I will adopt a scaling relation which associated the intensity of this field to the SFR density estimated in each annulus (similarly to what is assumed in Popping et al., 2019a; Krogager & Noterdaeme, 2020), while I will assume and universal value for the CR background, evolving with z. To describe the internal chemistry, and energetics of the molecular clouds I will adopt the DESPOTIC software by Mark Krumholz (available as a python/C++ package), which determines the equilibrium gas and dust temperatures in the molecular clouds, calculating time-dependent thermal evolution of clouds, and taking as input the

value of $G_{\rm UV}$, the CR primary ionization rate and the internal density profile of the molecular cloud. DESPOTIC will iteratively solve the equations for the gas and dust temperature and the carbon chemistry, by considering the chemical network implemented by Glover & Clark (2012) and all the available heating and cooling channels[2]. This study will consider at the same time a prescription to describe self-consistently the warm gas phase of the ISM and the clumpiness of molecular hydrogen in the ISM. This has not yet done in other semi-analytical models and I believe that it would lead to new important results on the cold gas cycle in galaxies, and in particular on the connection between DLA and galaxies at high redshift, the main motivation of my thesis work.

[2]The principal heating processes are: heating by the grain photoelectric effect, heating of the dust by the interstellar radiation field, and CR heating of the gas. The cooling is dominated by line cooling, as well as cooling of the dust by thermal emission.

Appendix A

Useful tables

TABLE of PHYSICAL AND ASTRONOMICAL CONSTANTS

Symbol	Definition	Value
c	Speed of light	2.998×10^8m s^{-1}
e	Elementary charge	1.602×10^{-19} C
m_n	Neutron rest mass	1.675×10^{-27} kg
m_p	Proton rest mass	1.673×10^{-27} kg
h	Planck's constant	6.626×10^{-34} J s
k_B	Boltzmann's constant	1.381×10^{-23} J K^{-1}
G	Gravitational constant	6.673×10^{-11} N m^2 kg^{-2}
σ	Stefan-Boltzmann constant	5.670×10^{-8} J m^{-2} K^{-4} s^{-1}
N_A	Avogadro constant	6.022 $\times 10^{23}$ mol^{-1}
α	Fine structure constant $(= 1/137.0)$	7.297 $\times 10^{-3}$
$M_\odot$	Solar Mass	1.989 $\times 10^{30}$ kg
$R_\odot$	Solar radius	6.96 $\times 10^8$ m
$L_\odot$	Solar luminosity	3.827 $\times 10^{26}$ J s^{-1}
1 l.y.	light year	9.461 $\times 10^{15}$ m
1 AU	Astronomical Unit	1.496 $\times 10^{11}$ m
1 pc	Parsec	3.086 $\times 10^{16}$ m
1 year		3.156 $\times 10^7$ s

Appendix B

Additional results for the GAEA-X17 model

B.1 Influence of different gas vertical density profiles on the estimated DLA properties

The distribution of HI gas detected through the 21 cm line is fairly flat and uniform (Leroy et al., 2008a), with a scalelength much larger than stellar disk one. The work by Narayan & Jog (2002) suggests that the vertical structure of the gaseous disk is sensitive to the gravity of all galactic components, i.e. stars, dark matter and gas. Under the assumption of an isothermal distribution, one expects that the gaseous/stellar vertical density profile is described by the function $sech^2$, as shown theoretically by Spitzer (1942) and confirmed by some observations (van der Kruit & Searle, 1982). However, more recent observational studies have found that the observed vertical distribution for gas and stars in galaxies is steeper than the one predicted by an isothermal distribution, and it is well-approximated by an exponential or a $sech$ function, especially close to the galactic mid-plane (Barteldrees & Dettmar, 1994; Rice et al., 1996)

We have considered the effect on the DLA column density distribution function (CDDF) of assuming a different vertical density profile for the model galaxies. We assume 4 different density profiles for the gas in the galactic

disc: the 'classic' double-exponential

$$\rho_{\rm CG}(r,z) = \rho_0 \; e^{-r/R_s} \; e^{-z/z_0}$$

and three additional profiles, described by the formula presented in van der Kruit & Freeman (2011):

$$\rho_{\rm CG}(r,z) = \rho_0 \; e^{-r/R_s} sech^{\frac{2}{n}}(\frac{n\,{\rm z}}{2z_0})$$

with $n = 1, 2$ and 4 respectively.

In Fig B.1 the predicted average CDDF in the redshift range 2izi3 is shown. The solid line refers to the exponential vertical profile, while the star-dashed, dot-dashed and dashed line refer to the function presented in van der Kruit & Freeman (2011), respectively with $n = 1,\ 2,\ 4$. Fig B.1 highlights that the 4 different density profiles lead to differences in the CDDF only in the high column-density regime. And for large n-values the class of functions presented by van der Kruit & Freeman (2011) give very similar results to the exponential density profile, for all the column densities considered.

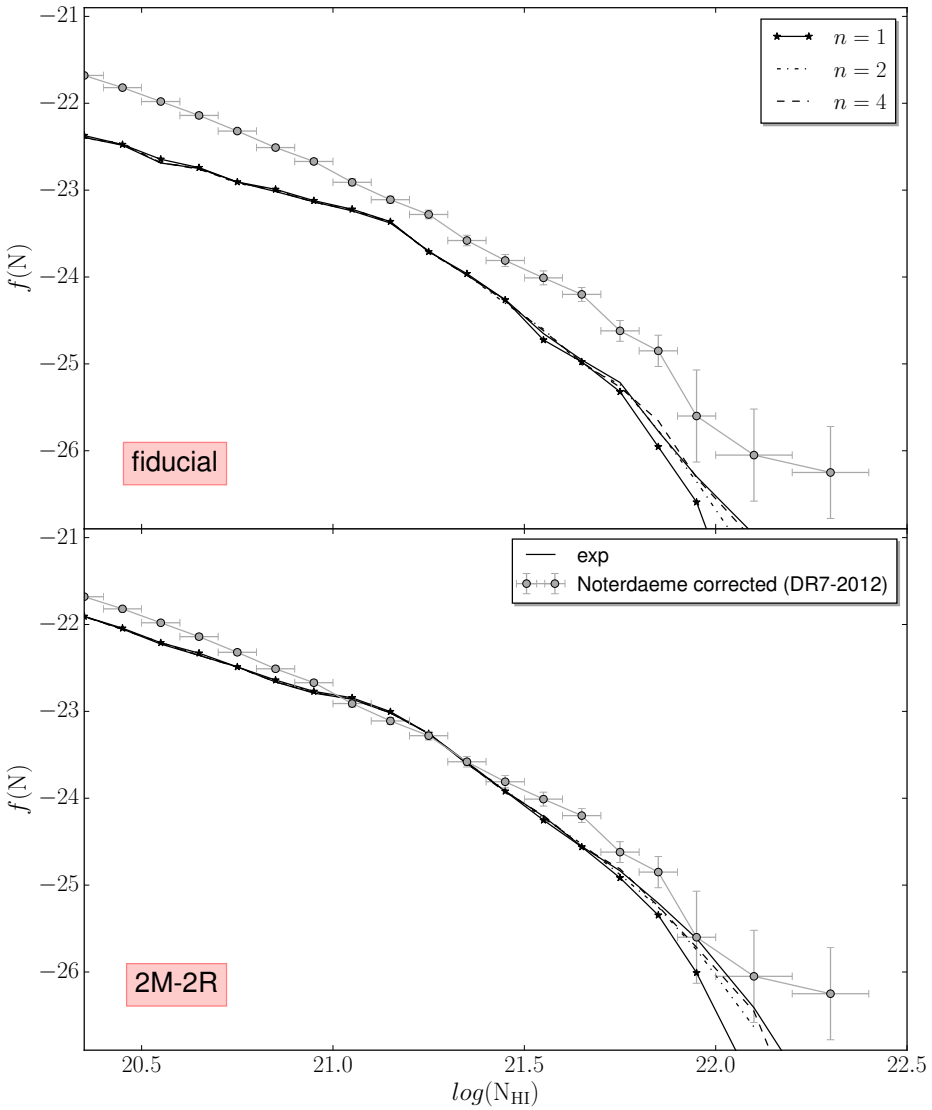

Fig. B.1: Average CDDF in the redshift range $2 < z < 3$, compared to the data (grey dots) by Noterdaeme et al. (2012). The top (bottom) panel shows the results from the fiducial ($2M - 2R$) model. The black lines describe the total CDDF while the other lines show the average contribution to the CDDF assuming different density profiles, as described in the legend.

List of Figures

References

Abel T., Wandelt B. D., 2002, MNRAS, 330, L53

Abel T., Bryan G. L., Norman M. L., 1998, Mem. Soc. Astron. Italiana, 69, 377

Accurso G., et al., 2017, MNRAS, 470, 4750

Agertz O., et al., 2007, MNRAS, 380, 963

Allen S. W., Schmidt R. W., Fabian A. C., Ebeling H., 2003, MNRAS, 342, 287

Altay G., Theuns T., Schaye J., Crighton N. H. M., Dalla Vecchia C., 2011, ApJ, 737, L37

Altay G., Theuns T., Schaye J., Booth C. M., Dalla Vecchia C., 2013, MNRAS, 436, 2689

Andreani P., Miyamoto Y., Kaneko H., Boselli A., Tatematsu K., Sorai K., Vio R., 2020, A&A, 643, L11

Aracil B., Petitjean P., Pichon C., Bergeron J., 2004, A&A, 419, 811

Aravena M., et al., 2012, Unveiling the cold molecular gas in the brightest and most distant galaxies discovered with the SPT, ATNF Proposal

Aravena M., et al., 2016, MNRAS, 457, 4406

Arinyo-i-Prats A., Mas-Ribas L., Miralda-Escudé J., Pérez-Ràfols I., Noterdaeme P., 2018, MNRAS, 481, 3921

Arth A., Dolag K., Beck A. M., Petkova M., Lesch H., 2014, arXiv e-prints, p. arXiv:1412.6533

Ascasibar Y., Yepes G., Gottlöber S., Müller V., 2002, A&A, 387, 396

Asplund M., Grevesse N., Sauval A. J., Scott P., 2009, ARA&A, 47, 481

Aubourg É., et al., 2015, Phys. Rev. D, 92, 123516

Bacon R., et al., 2021, A&A, 647, A107

Bahcall J. N., Salpeter E. E., 1965, ApJ, 142, 1677

Balashev S. A., Noterdaeme P., 2018, MNRAS, 478, L7

Baldry I. K., et al., 2012, MNRAS, 421, 621

Barnes L. A., Haehnelt M. G., 2009, MNRAS, 397, 511

Barnes L. A., Garel T., Kacprzak G. G., 2014, PASP, 126, 969

Barrow J. D., Turner M. S., 1981, Nature, 292, 35

Barrow K. S. S., Wise J. H., Norman M. L., O'Shea B. W., Xu H., 2017, MNRAS, 469, 4863

Barteldrees A., Dettmar R. J., 1994, A&AS, 103, 475

Baugh C. M., 2006, Reports on Progress in Physics, 69, 3101

Baugh C. M., Lacey C. G., Frenk C. S., Benson A. J., Cole S., Granato G. L., Silva L., Bressan A., 2004, New A Rev., 48, 1239

Becker G. D., Bolton J. S., 2013, MNRAS, 436, 1023

Becker R. H., et al., 2001, AJ, 122, 2850

Becker G. D., Rauch M., Sargent W. L. W., 2007, ApJ, 662, 72

Becker G. D., Bolton J. S., Haehnelt M. G., Sargent W. L. W., 2011, MNRAS, 410, 1096

Becker G. D., Bolton J. S., Madau P., Pettini M., Ryan-Weber E. V., Venemans B. P., 2015, MNRAS, 447, 3402

Behroozi P. S., Conroy C., Wechsler R. H., 2010, ApJ, 717, 379

Bennett C. L., et al., 2013, ApJS, 208, 20

Benson A. J., 2010, Phys. Rep., 495, 33

Berlind A. A., Weinberg D. H., 2002, ApJ, 575, 587

Berry M., Somerville R. S., Haas M. R., Gawiser E., Maller A., Popping G., Trager S. C., 2014, MNRAS, 441, 939

Berry M., Somerville R. S., Gawiser E., Maller A. H., Popping G., Trager S. C., 2016, MNRAS, 458, 531

Bialy S., Sternberg A., 2019, ApJ, 881, 160

Bigiel F., Leroy A., Walter F., Brinks E., de Blok W. J. G., Madore B., Thornley M. D., 2008, AJ, 136, 2846

Bigiel F., et al., 2011, ApJ, 730, L13

Bird S., Vogelsberger M., Haehnelt M., Sijacki D., Genel S., Torrey P., Springel V., Hernquist L., 2014, MNRAS, 445, 2313

Bird S., Haehnelt M., Neeleman M., Genel S., Vogelsberger M., Hernquist L., 2015, MNRAS, 447, 1834

Blitz L., Rosolowsky E., 2006, ApJ, 650, 933

Blumenthal G. R., Faber S. M., Primack J. R., Rees M. J., 1984, Nature, 311, 517

Boera E., Murphy M. T., Becker G. D., Bolton J. S., 2014, MNRAS, 441, 1916

Boksenberg A., Carswell R. F., Smith M. G., Whelan J. A. J., 1978, MNRAS, 184, 773

Bolatto A. D., Wolfire M., Leroy A. K., 2013, ARA&A, 51, 207

Bolton J. S., Becker G. D., Haehnelt M. G., Viel M., 2014, MNRAS, 438, 2499

Boogaard L. A., et al., 2019, ApJ, 882, 140

Booth C. M., Schaye J., 2009, MNRAS, 398, 53

Boselli A., Cortese L., Boquien M., 2014, A&A, 564, A65

Bouché N., Hohensee W., Vargas R., Kacprzak G. G., Martin C. L., Cooke J., Churchill C. W., 2012, MNRAS, 426, 801

Boylan-Kolchin M., Springel V., White S. D. M., Jenkins A., Lemson G., 2009, MNRAS, 398, 1150

Brook C. B., Kawata D., Gibson B. K., Flynn C., 2004, MNRAS, 349, 52

Bryan G. L., Machacek M. E., 2000, ApJ, 534, 57

Bryan G. L., et al., 2014, ApJS, 211, 19

Burbidge E. M., Lynds C. R., Burbidge G. R., 1966, ApJ, 144, 447

Cantalupo S., Porciani C., 2011, MNRAS, 411, 1678

Carr B. J., Bond J. R., Arnett W. D., 1984, ApJ, 277, 445

Carswell R. F., Webb J. K., 2014, VPFIT: Voigt profile fitting program (ascl:1408.015)

Carton D., et al., 2018, MNRAS, 478, 4293

Catinella B., et al., 2010, MNRAS, 403, 683

Catinella B., et al., 2013, MNRAS, 436, 34

Catinella B., et al., 2018, MNRAS, 476, 875

Cen R., 2010, ApJ, 725, 115

Cen R., 2012, ApJ, 748, 121

Cen R., Ostriker J. P., 1992, ApJ, 399, L113

Cen R., Ostriker J. P., 1999, ApJ, 519, L109

Cen R., Ostriker J. P., Prochaska J. X., Wolfe A. M., 2003, ApJ, 598, 741

Chen H.-W., Lanzetta K. M., 2003, ApJ, 597, 706

Choi E., Brennan R., Somerville R. S., Ostriker J. P., Hirschmann M., Naab T., 2020, ApJ, 904, 8

Christensen C. R., Governato F., Quinn T., Brooks A. M., Shen S., McCleary J., Fisher D. B., Wadsley J., 2014a, MNRAS, 440, 2843

Christensen L., Møller P., Fynbo J. P. U., Zafar T., 2014b, MNRAS, 445, 225

Christensen L., Møller P., Rhodin N. H. P., Heintz K. E., Fynbo J. P. U., 2019, MNRAS, 489, 2270

Cole S., 1991, ApJ, 367, 45

Cole S., Lacey C. G., Baugh C. M., Frenk C. S., 2000, MNRAS, 319, 168

Cole S., et al., 2005, MNRAS, 362, 505

Conroy C., Wechsler R. H., Kravtsov A. V., 2006, ApJ, 647, 201

Cooke J., Wolfe A. M., Gawiser E., Prochaska J. X., 2006, ApJ, 652, 994

Couchman H., Pearce F., Thomas P., 2011, Hydra: A Parallel Adaptive Grid Code (ascl:1103.010)

Crain R. A., et al., 2015, MNRAS, 450, 1937

Crain R. A., et al., 2017, MNRAS, 464, 4204

Crighton N. H. M., et al., 2015, MNRAS, 452, 217

Croton D. J., et al., 2006, MNRAS, 365, 11

D'Odorico V., et al., 2016, MNRAS, 463, 2690

Daddi E., et al., 2010, ApJ, 713, 686

Dalla Vecchia C., Schaye J., 2008, MNRAS, 387, 1431

Davé R., Hellsten U., Hernquist L., Katz N., Weinberg D. H., 1998, ApJ, 509, 661

Davé R., Katz N., Oppenheimer B. D., Kollmeier J. A., Weinberg D. H., 2013, MNRAS, 434, 2645

Davé R., Anglés-Alcázar D., Narayanan D., Li Q., Rafieferantsoa M. H., Appleby S., 2019, MNRAS, 486, 2827

Davis T. A., et al., 2011, MNRAS, 417, 882

De Cia A., Ledoux C., Petitjean P., Savaglio S., 2018, Astronomy & Astrophysics, 611, A76

De Lucia G., Blaizot J., 2007, MNRAS, 375, 2

De Lucia G., Tornatore L., Frenk C. S., Helmi A., Navarro J. F., White S. D. M., 2014, MNRAS, 445, 970

Debuhr J., Quataert E., Ma C.-P., 2011, MNRAS, 412, 1341

Decarli R., et al., 2016, ApJ, 833, 70

Decarli R., et al., 2019, ApJ, 882, 138

Dekel A., Silk J., 1986, ApJ, 303, 39

Di Gioia S., Cristiani S., De Lucia G., Xie L., 2020, MNRAS, 497, 2469

Di Matteo T., Springel V., Hernquist L., 2005, Nature, 433, 604

Diemer B., et al., 2018, ApJS, 238, 33

Diemer B., et al., 2019, MNRAS, 487, 1529

Djorgovski S. G., Castro S., Stern D., Mahabal A. A., 2001, ApJ, 560, L5

Dolag K., Stasyszyn F., 2009, MNRAS, 398, 1678

Dolag K., Jubelgas M., Springel V., Borgani S., Rasia E., 2004, ApJ, 606, L97

Dubois Y., Teyssier R., 2008a, in Knapen J. H., Mahoney T. J., Vazdekis A., eds, Astronomical Society of the Pacific Conference Series Vol. 390, Pathways Through an Eclectic Universe. p. 388

Dubois Y., Teyssier R., 2008b, A&A, 482, L13

Dubois Y., Devriendt J., Slyz A., Teyssier R., 2012, MNRAS, 420, 2662

Durkalec A., et al., 2015, A&A, 576, L7

Dutta R., et al., 2020, MNRAS, 499, 5022

Efstathiou G., 1992, MNRAS, 256, 43P

Efstathiou G., Eastwood J. W., 1981, MNRAS, 194, 503

Eisenstein D. J., et al., 2011, AJ, 142, 72

Elmegreen B. G., 1993, ApJ, 411, 170

Elmegreen B. G., Elmegreen D. M., Tompkins B., Jenks L. G., 2017, ApJ, 847, 14

Erkal D., Gnedin N. Y., Kravtsov A. V., 2012, ApJ, 761, 54

Evrard A. E., 1988, MNRAS, 235, 911

Fan X., et al., 2006, AJ, 132, 117

Faucher-Giguère C.-A., Kereš D., 2011, MNRAS, 412, L118

Faucher-Giguère C.-A., Lidz A., Zaldarriaga M., Hernquist L., 2009, ApJ, 703, 1416

Feldmann R., Gnedin N. Y., Kravtsov A. V., 2011, ApJ, 732, 115

Feldmann R., Hernandez J., Gnedin N. Y., 2012, ApJ, 761, 167

Ferrara A., Pettini M., Shchekinov Y., 2000, MNRAS, 319, 539

Few C. G., Courty S., Gibson B. K., Kawata D., Calura F., Teyssier R., 2012, MNRAS, 424, L11

Fletcher T. J., Saintonge A., Soares P. S., Pontzen A., 2021, MNRAS, 501, 411

Font-Ribera A., et al., 2012, Journal of Cosmology and Astroparticle Physics, 2012

Fontana A., Ballester P., 1995, The Messenger, 80, 37

Fontana A., et al., 2006, A&A, 459, 745

Frenk C. S., et al., 1999, ApJ, 525, 554

Fryxell B., et al., 2000, ApJS, 131, 273

Fu J., Guo Q., Kauffmann G., Krumholz M. R., 2010, MNRAS, 409, 515

Fumagalli M., O'Meara J. M., Prochaska J. X., Kanekar N., 2010, MNRAS, 408, 362

Fumagalli M., Prochaska J. X., Kasen D., Dekel A., Ceverino D., Primack J. R., 2011, MNRAS, 418, 1796

Fumagalli M., O'Meara J. M., Prochaska J. X., Rafelski M., Kanekar N., 2015, MNRAS, 446, 3178

Furlanetto S. R., Oh S. P., 2009, ApJ, 701, 94

Fynbo J. U., Møller P., Warren S. J., 1999, MNRAS, 305, 849

Fynbo J. P. U., Prochaska J. X., Sommer-Larsen J., Dessauges-Zavadsky M., Møller P., 2008, ApJ, 683, 321

Fynbo J. P. U., et al., 2010, MNRAS, 408, 2128

Fynbo J. P. U., et al., 2011, MNRAS, 413, 2481

Fynbo J. P. U., Krogager J. K., Venemans B., Noterdaeme P., Vestergaard M., Møller P., Ledoux C., Geier S., 2013, ApJS, 204, 6

Gaburov E., Nitadori K., 2011, MNRAS, 414, 129

Gardner J. P., Katz N., Hernquist L., Weinberg D. H., 1997, ApJ, 484, 31

Gardner J. P., Katz N., Hernquist L., Weinberg D. H., 2001, ApJ, 559, 131

Garnavich P. M., et al., 1998, ApJ, 509, 74

Garratt-Smithson L., Power C., Lagos C. d. P., Stevens A. R. H., Allison J. R., Sadler E. M., 2021, MNRAS, 501, 4396

Genzel R., et al., 2010, MNRAS, 407, 2091

Genzel R., et al., 2012, ApJ, 746, 69

Genzel R., et al., 2015, ApJ, 800, 20

Gerin M., et al., 2015, A&A, 573, A30

Gerritsen J. P. E., Icke V., 1997, A&A, 325, 972

Gingold R. A., Monaghan J. J., 1977, MNRAS, 181, 375

Giovanelli R., et al., 2005, AJ, 130, 2598

Glover S. C. O., Clark P. C., 2012, MNRAS, 421, 116

Gnedin N. Y., 2000, ApJ, 542, 535

Gnedin N. Y., Draine B. T., 2014, ApJ, 795, 37

Gnedin N. Y., Kravtsov A. V., 2010, The Astrophysical Journal, 714, 287

Gnedin N. Y., Kravtsov A. V., 2011, ApJ, 728, 88

Gnedin N. Y., Ostriker J. P., 1997, ApJ, 486, 581

Greig B., Mesinger A., 2017, MNRAS, 465, 4838

Greig B., Trott C. M., Barry N., Mutch S. J., Pindor B., Webster R. L., Wyithe J. S. B., 2021, MNRAS, 500, 5322

Gunn J. E., Peterson B. A., 1965, ApJ, 142, 1633

Guo Q., White S., Angulo R. E., Henriques B., Lemson G., Boylan-Kolchin M., Thomas P., Short C., 2013, MNRAS, 428, 1351

Gupta N., et al., 2021a, arXiv e-prints, p. arXiv:2103.09437

Gupta N., et al., 2021b, ApJ, 907, 11

Guzzo L., et al., 2008, Nature, 451, 541

Haardt F., Madau P., 1996, ApJ, 461, 20

Haehnelt M. G., Steinmetz M., Rauch M., 1998, ApJ, 495, 647

Haiman Z., Loeb A., 1997, ApJ, 483, 21

Hamanowicz A., et al., 2020, MNRAS, 492, 2347

Hassan S., Finlator K., Davé R., Churchill C. W., Prochaska J. X., 2019, arXiv e-prints, p. arXiv:1910.07541

Hassan S., Finlator K., Davé R., Churchill C. W., Prochaska J. X., 2020, MNRAS, 492, 2835

Hayward C. C., Torrey P., Springel V., Hernquist L., Vogelsberger M., 2014, MNRAS, 442, 1992

Heitsch F., Naab T., Walch S., 2011, MNRAS, 415, 271

Henriques B. M. B., Yates R. M., Fu J., Guo Q., Kauffmann G., Srisawat C., Thomas P. A., White S. D. M., 2020, MNRAS, 491, 5795

Hernquist L., Katz N., 1989, ApJS, 70, 419

Hill A. S., Mac Low M.-M., Gatto A., Ibáñez-Mejía J. C., 2018, ApJ, 862, 55

Hinshaw G., et al., 2013, ApJS, 208, 19

Hirschmann M., De Lucia G., Fontanot F., 2016, MNRAS, 461, 1760

Hoeft M., Yepes G., Gottlöber S., Springel V., 2006, MNRAS, 371, 401

Hopkins P. F., 2014, GIZMO: Multi-method magneto-hydrodynamics+gravity code (ascl:1410.003)

Hopkins P. F., 2015, MNRAS, 450, 53

Hopkins P. F., Kereš D., Oñorbe J., Faucher-Giguère C.-A., Quataert E., Murray N., Bullock J. S., 2014, MNRAS, 445, 581

Hopkins P. F., et al., 2018, MNRAS, 480, 800

Hu W., 2000, ApJ, 529, 12

Hu E. M., Kim T.-S., Cowie L. L., Songaila A., Rauch M., 1995, AJ, 110, 1526

Hu C.-Y., Naab T., Walch S., Moster B. P., Oser L., 2014, MNRAS, 443, 1173

Hubber D. A., Falle S. A. E. G., Goodwin S. P., 2013, MNRAS, 432, 711

Hui L., Gnedin N. Y., 1997, MNRAS, 292, 27

Jubelgas M., Springel V., Dolag K., 2004, MNRAS, 351, 423

Jubelgas M., Springel V., Enßlin T., Pfrommer C., 2008, A&A, 481, 33

Kacprzak G. G., Churchill C. W., Nielsen N. M., 2012, ApJ, 760, L7

Kacprzak G. G., Muzahid S., Churchill C. W., Nielsen N. M., Charlton J. C., 2015, ApJ, 815, 22

Kalberla P. M. W., Kerp J., 2009, ARA&A, 47, 27

Kanekar N., et al., 2014, MNRAS, 438, 2131

Katz N., Weinberg D. H., Hernquist L., 1996, ApJS, 105, 19

Katz H., Kimm T., Haehnelt M. G., Sijacki D., Rosdahl J., Blaizot J., 2019, MNRAS, 483, 1029

Kauffmann G., Colberg J. M., Diaferio A., White S. D. M., 1999, MNRAS, 303, 188

Kauffmann G., Nelson D., Borthakur S., Heckman T., Hernquist L., Marinacci F., Pakmor R., Pillepich A., 2019, MNRAS, 486, 4686

Kawata D., Gibson B. K., 2003, MNRAS, 340, 908

Kawata D., Gibson B. K., 2005, MNRAS, 358, L16

Keenan R. P., Marrone D. P., Keating G. K., 2020, ApJ, 904, 127

Keres D., Yun M. S., Young J. S., 2003, ApJ, 582, 659

Kereš D., Katz N., Weinberg D. H., Davé R., 2005, MNRAS, 363, 2

Khare P., Vanden Berk D., York D. G., Lundgren B., Kulkarni V. P., 2012, MNRAS, 419, 1028

Kim T. S., Carswell R. F., Cristiani S., D'Odorico S., Giallongo E., 2002a, MNRAS, 335, 555

Kim T. S., Cristiani S., D'Odorico S., 2002b, A&A, 383, 747

Kim J.-h., et al., 2014, ApJS, 210, 14

Kim H.-S., Wyithe J. S. B., Power C., Park J., Lagos C. d. P., Baugh C. M., 2015, MNRAS, 453, 2315

Kirkman D., Tytler D., 1997, ApJ, 484, 672

Klitsch A., et al., 2019, MNRAS, 490, 1220

Kobayashi C., 2004, MNRAS, 347, 740

Kohler K., Gnedin N. Y., Hamilton A. J. S., 2007, ApJ, 657, 15

Kravtsov A. V., Klypin A. A., Khokhlov A. M., 1997, ApJS, 111, 73

Kravtsov A. V., Gnedin O. Y., Klypin A. A., 2004, ApJ, 609, 482

Kregel M., van der Kruit P. C., de Grijs R., 2002, MNRAS, 334, 646

Krogager J.-K., Noterdaeme P., 2020, A&A, 644, L6

Krogager J. K., et al., 2015, ApJS, 217, 5

Krogager J. K., et al., 2016, ApJ, 832, 49

Krogager J. K., Møller P., Fynbo J. P., Noterdaeme P., 2017, Monthly Notices of the Royal Astronomical Society, 469, 2959

Krumholz M. R., 2012, ApJ, 759, 9

Krumholz M. R., McKee C. F., Tumlinson J., 2008, ApJ, 689, 865

Krumholz M. R., McKee C. F., Tumlinson J., 2009a, ApJ, 693, 216

Krumholz M. R., McKee C. F., Tumlinson J., 2009b, ApJ, 699, 850

Kulkarni V. P., Khare P., Péroux C., York D. G., Lauroesch J. T., Meiring J. D., 2007, ApJ, 661, 88

Kulkarni V. P., Khare P., Som D., Meiring J., York D. G., Péroux C., Lauroesch J. T., 2010, New A, 15, 735

Kurosawa R., Proga D., Nagamine K., 2009, ApJ, 707, 823

Lacey C., Silk J., 1991, ApJ, 381, 14

Lagos C. D. P., Baugh C. M., Lacey C. G., Benson A. J., Kim H.-S., Power C., 2011, MNRAS, 418, 1649

Lagos C. D. P., Baugh C. M., Zwaan M. A., Lacey C. G., Gonzalez-Perez V., Power C., Swinbank A. M., van Kampen E., 2014, MNRAS, 440, 920

Lagos C. d. P., Tobar R. J., Robotham A. S. G., Obreschkow D., Mitchell P. D., Power C., Elahi P. J., 2018, MNRAS, 481, 3573

Lah P., et al., 2007, MNRAS, 376, 1357

Lenkić L., et al., 2020, AJ, 159, 190

Leroy A. K., Walter F., Brinks E., Bigiel F., de Blok W. J. G., Madore B., Thornley M. D., 2008a, AJ, 136, 2782

Leroy A. K., Walter F., Brinks E., Bigiel F., de Blok W. J. G., Madore B., Thornley M. D., 2008b, AJ, 136, 2782

Leroy A. K., et al., 2013, AJ, 146, 19

Li Y.-S., De Lucia G., Helmi A., 2010, MNRAS, 401, 2036

Lia C., Portinari L., Carraro G., 2002, MNRAS, 330, 821

Lovell C. C., Vijayan A. P., Thomas P. A., Wilkins S. M., Barnes D. J., Irodotou D., Roper W., 2021, MNRAS, 500, 2127

Ludlow A. D., et al., 2013, MNRAS, 432, 1103

Ma X., 2018, PhD thesis, TAPIR, MC 350-17, California Institute of Technology, Pasadena, CA 91125, USA

Ma X., Hopkins P. F., Feldmann R., Torrey P., Faucher-Giguère C.-A., Kereš D., 2017, MNRAS, 466, 4780

Ma X., et al., 2018, MNRAS, 478, 1694

Ma X., Quataert E., Wetzel A., Hopkins P. F., Faucher-Giguère C.-A., Kereš D., 2020, MNRAS, 498, 2001

Mac Low M.-M., Ferrara A., 1999, ApJ, 513, 142

Macciò A. V., Kang X., Fontanot F., Somerville R. S., Koposov S., Monaco P., 2010, MNRAS, 402, 1995

Mackenzie R., et al., 2019, MNRAS, 487, 5070

Madau P., Ferrara A., Rees M. J., 2001, ApJ, 555, 92

Mandelbaum R., 2018, ARA&A, 56, 393

Martin A. M., Papastergis E., Giovanelli R., Haynes M. P., Springob C. M., Stierwalt S., 2010, ApJ, 723, 1359

Martin P. G., Blagrave K. P. M., Lockman F. J., Pinheiro Gonçalves D., Boothroyd A. I., Joncas G., Miville-Deschênes M. A., Stephan G., 2015, ApJ, 809, 153

Mather J. C., et al., 1990, ApJ, 354, L37

McCrea W. H., 1972, Planet. Space Sci., 20, 1796

McDonald P., Miralda-Escudé J., 2001, ApJ, 549, L11

McDonald P., Miralda-Escudé J., Rauch M., Sargent W. L. W., Barlow T. A., Cen R., 2001, ApJ, 562, 52

McLeod D. J., McLure R. J., Dunlop J. S., Cullen F., Carnall A. C., Duncan K., 2021, MNRAS, 503, 4413

McQuinn M., Oh S. P., Faucher-Giguère C.-A., 2011, ApJ, 743, 82

Meiksin A., White M., 2004, MNRAS, 350, 1107

Menon H., Wesolowski L., Zheng G., Jetley P., Kale L., Quinn T., Governato F., 2015, Computational Astrophysics and Cosmology, 2, 1

Meyer M. J., et al., 2004, MNRAS, 350, 1195

Mihos J. C., Hernquist L., 1994, ApJ, 437, 611

Miralda-Escudé J., 2003, ApJ, 597, 66

Møller P., Christensen L., 2020, MNRAS, 492, 4805

Møller P., Warren S. J., 1998, MNRAS, 299, 661

Møller P., Warren S. J., Fall S. M., Fynbo J. U., Jakobsen P., 2002, ApJ, 574, 51

Mosconi M. B., Tissera P. B., Lambas D. G., Cora S. A., 2001, MNRAS, 325, 34

Moster B. P., Naab T., White S. D. M., 2013, MNRAS, 428, 3121

Moustakas J., et al., 2013, ApJ, 767, 50

Naab T., Ostriker J. P., 2017, ARA&A, 55, 59

Nagamine K., Springel V., Hernquist L., 2004, MNRAS, 348, 421

Nagamine K., Wolfe A. M., Hernquist L., Springel V., 2007, ApJ, 660, 945

Narayan C. A., Jog C. J., 2002, A&A, 394, 89

Nath B. B., Trentham N., 1997, MNRAS, 291, 505

Navarro J. F., White S. D. M., 1993, MNRAS, 265, 271

Navarro J. F., Frenk C. S., White S. D. M., 1995, MNRAS, 275, 56

Navarro J. F., Frenk C. S., White S. D. M., 1996, ApJ, 462, 563

Neeleman M., Wolfe A. M., Prochaska J. X., Rafelski M., 2013, ApJ, 769, 54

Neeleman M., Kanekar N., Prochaska J. X., Rafelski M. A., Carilli C. L., 2019, ApJ, 870, L19

Neistein E., Dekel A., 2008, MNRAS, 383, 615

Noh Y., McQuinn M., 2014, MNRAS, 444, 503

Noterdaeme P., et al., 2012, A&A, 547, L1

Noterdaeme P., Srianand R., Rahmani H., Petitjean P., Pâris I., Ledoux C., Gupta N., López S., 2015, A&A, 577, A24

O'Meara J. M., Prochaska J. X., Worseck G., Chen H.-W., Madau P., 2013, ApJ, 765, 137

Obreschkow D., Rawlings S., 2009, MNRAS, 394, 1857

Okamoto T., Nemmen R. S., Bower R. G., 2008a, MNRAS, 385, 161

Okamoto T., Gao L., Theuns T., 2008b, MNRAS, 390, 920

Okamoto T., Frenk C. S., Jenkins A., Theuns T., 2010, MNRAS, 406, 208

Oppenheimer B. D., Davé R., 2006, MNRAS, 373, 1265

Oppenheimer B. D., Davé R., 2008, MNRAS, 387, 577

Ostriker J. P., Gnedin N. Y., 1996, ApJ, 472, L63

Ostriker J. P., Vishniac E. T., 1986, ApJ, 306, L51

Padmanabhan H., Refregier A., Amara A., 2017, MNRAS, 469, 2323

Pakmor R., Springel V., 2013, MNRAS, 432, 176

Pakmor R., Bauer A., Springel V., 2011, MNRAS, 418, 1392

Pandya V., et al., 2021, arXiv e-prints, p. arXiv:2103.06891

Parkinson H., Cole S., Helly J., 2008, MNRAS, 383, 557

Peeples M. S., Werk J. K., Tumlinson J., Oppenheimer B. D., Prochaska J. X., Katz N., Weinberg D. H., 2014, ApJ, 786, 54

Pehlivan Rhodin N. H., Agertz O., Christensen L., Renaud F., Uldall Fynbo J. P., 2019, arXiv e-prints, p. arXiv:1901.10777

Pentericci L., et al., 2002, AJ, 123, 2151

Pérez-González P. G., et al., 2008, ApJ, 675, 234

Pérez-Ràfols I., et al., 2018a, MNRAS, 473, 3019

Pérez-Ràfols I., Miralda-Escudé J., Arinyo-i-Prats A., Font-Ribera A., Mas-Ribas L., 2018b, MNRAS, 480, 4702

Perlmutter S., et al., 1999, ApJ, 517, 565

Péroux C., Howk J. C., 2020, ARA&A, 58, 363

Péroux C., McMahon R. G., Storrie-Lombardi L. J., Irwin M. J., 2003, MNRAS, 346, 1103

Péroux C., Bouché N., Kulkarni V. P., York D. G., Vladilo G., 2011, MNRAS, 410, 2237

Péroux C., et al., 2019, Monthly Notices of the Royal Astronomical Society, 485, 1595

Petkova M., Springel V., 2011a, MNRAS, 412, 935

Petkova M., Springel V., 2011b, MNRAS, 415, 3731

Pillepich A., et al., 2018, MNRAS, 475, 648

Piontek F., Steinmetz M., 2011, MNRAS, 410, 2625

Planck Collaboration et al., 2016, A&A, 594, A1

Planck Collaboration et al., 2018, arXiv e-prints, p. arXiv:1807.06209

Pontzen A., et al., 2008, MNRAS, 390, 1349

Popping G., Somerville R. S., Trager S. C., 2014, MNRAS, 442, 2398

Popping G., et al., 2017, A&A, 602, A11

Popping G., Narayanan D., Somerville R. S., Faisst A. L., Krumholz M. R., 2019a, MNRAS, 482, 4906

Popping G., et al., 2019b, ApJ, 882, 137

Power C., Baugh C. M., Lacey C. G., 2010, MNRAS, 406, 43

Price D. J., Federrath C., 2010, in Pogorelov N. V., Audit E., Zank G. P., eds, Astronomical Society of the Pacific Conference Series Vol. 429, Numerical Modeling of Space Plasma Flows, Astronum-2009. p. 274 (arXiv:0910.0285)

Pritchard J. R., Furlanetto S. R., 2007, MNRAS, 376, 1680

Prochaska J. X., Wolfe A. M., 1997, ApJ, 487, 73

Prochaska J. X., Wolfe A. M., 2009, The Astrophysical Journal, 696, 1543

Prochaska J. X., Castro S., Djorgovski S. G., 2003, ApJS, 148, 317

Prochaska J. X., Herbert-Fort S., Wolfe A. M., 2005, The Astrophysical Journal, 635, 123

Prochaska J. X., O'Meara J. M., Fumagalli M., Bernstein R. A., Burles S. M., 2015, ApJS, 221, 2

Rafelski M., Wolfe A. M., Prochaska J. X., Neeleman M., Mendez A. J., 2012, ApJ, 755, 89

Rahmani H., et al., 2016, MNRAS, 463, 980

Rahmati A., Schaye J., 2014, MNRAS, 438, 529

Rahmati A., Pawlik A. H., Raičević M., Schaye J., 2013, MNRAS, 430, 2427

Rahmati A., Schaye J., Bower R. G., Crain R. A., Furlong M., Schaller M., Theuns T., 2015, MNRAS, 452, 2034

Rao S. M., Turnshek D. A., Nestor D. B., 2006, ApJ, 636, 610

Rao S. M., Belfort-Mihalyi M., Turnshek D. A., Monier E. M., Nestor D. B., Quider A., 2011a, MNRAS, 416, 1215

Rao S. M., Belfort-Mihalyi M., Turnshek D. A., Monier E. M., Nestor D. B., Quider A., 2011b, MNRAS, 416, 1215

Razoumov A. O., 2009, The Astrophysical Journal, 707, 738

Read J. I., Hayfield T., 2012, MNRAS, 422, 3037

Rees M. J., 1986, MNRAS, 218, 25P

Rhodin N. H. P., Christensen L., Moller P., Zafar T., Fynbo J. P. U., 2018, A&A, 618, A129

Rhodin N. H. P., Agertz O., Christensen L., Renaud F., Fynbo J. P. U., 2019, MNRAS, 488, 3634

Ribaudo J., Lehner N., Howk J. C., 2011, ApJ, 736, 42

Rice W., Merrill K. M., Gatley I., Gillett F. C., 1996, AJ, 112, 114

Richings A. J., Schaye J., 2016, MNRAS, 458, 270

Ricotti M., Gnedin N. Y., Shull J. M., 2000, ApJ, 534, 41

Riechers D. A., et al., 2019, ApJ, 872, 7

Riechers D. A., et al., 2020, ApJ, 896, L21

Riess A. G., et al., 1998, AJ, 116, 1009

Romano M., Grazian A., Giallongo E., Cristiani S., Fontanot F., Boutsia K., Fiore F., Menci N., 2019, A&A, 632, A45

Rosdahl J., et al., 2018, MNRAS, 479, 994

Rosenberg J. L., Schneider S. E., 2003, ApJ, 585, 256

Roukema B. F., Quinn P. J., Peterson B. A., Rocca-Volmerange B., 1997, MNRAS, 292, 835

Rubin K. H. R., Hennawi J. F., Prochaska J. X., Simcoe R. A., Myers A., Lau M. W., 2015, ApJ, 808, 38

Rudie G. C., Steidel C. C., Shapley A. E., Pettini M., 2013, ApJ, 769, 146

Saintonge A., et al., 2011, MNRAS, 415, 61

Saintonge A., et al., 2017, ApJS, 233, 22

Saitta F., D'Odorico V., Bruscoli M., Cristiani S., Monaco P., Viel M., 2008, MNRAS, 385, 519

Sargent W. L. W., Young P. J., Boksenberg A., Tytler D., 1980, ApJS, 42, 41

Sargent W. L. W., Steidel C. C., Boksenberg A., 1989, ApJS, 69, 703

Scannapieco C., Tissera P. B., White S. D. M., Springel V., 2005, MNRAS, 364, 552

Scannapieco C., et al., 2012, MNRAS, 423, 1726

Schaller M., Dalla Vecchia C., Schaye J., Bower R. G., Theuns T., Crain R. A., Furlong M., McCarthy I. G., 2015, MNRAS, 454, 2277

Schaye J., 2001, ApJ, 559, 507

Schaye J., Dalla Vecchia C., 2008, MNRAS, 383, 1210

Schaye J., Theuns T., Leonard A., Efstathiou G., 1999, MNRAS, 310, 57

Schaye J., Theuns T., Rauch M., Efstathiou G., Sargent W. L. W., 2000, MNRAS, 318, 817

Schaye J., et al., 2015, MNRAS, 446, 521

Schneider D. P., et al., 2010, AJ, 139, 2360

Seljak U., Slosar A., McDonald P., 2006, J. Cosmology Astropart. Phys., 2006, 014

Sijacki D., Springel V., 2006, MNRAS, 366, 397

Sijacki D., Springel V., Di Matteo T., Hernquist L., 2007, MNRAS, 380, 877

Sijacki D., Vogelsberger M., Kereš D., Springel V., Hernquist L., 2012, MNRAS, 424, 2999

Simpson C. M., Bryan G. L., Hummels C., Ostriker J. P., 2015, ApJ, 809, 69

Slosar A., et al., 2013, Journal of Cosmology and Astroparticle Physics, 2013, 026

Smoot G. F., 1999, in Maiani L., Melchiorri F., Vittorio N., eds, American Institute of Physics Conference Series Vol. 476, 3K cosmology. pp 1–10 (arXiv:astro-ph/9902027), doi:10.1063/1.59326

Sobacchi E., Mesinger A., 2013, MNRAS, 432, L51

Somerville R. S., Davé R., 2015, ARA&A, 53, 51

Somerville R. S., Primack J. R., 1999, MNRAS, 310, 1087

Somerville R. S., Primack J. R., Faber S. M., 2001, MNRAS, 320, 504

Somerville R. S., et al., 2008, The Astrophysical Journal, 672, 776

Somerville R. S., Popping G., Trager S. C., 2015, MNRAS, 453, 4337

Sommer-Larsen J., Götz M., Portinari L., 2003, ApJ, 596, 47

Spinelli M., Zoldan A., De Lucia G., Xie L., Viel M., 2020, MNRAS, 493, 5434

Spitzer Lyman J., 1942, ApJ, 95, 329

Springel V., 2010a, ARA&A, 48, 391

Springel V., 2010b, MNRAS, 401, 791

Springel V., Hernquist L., 2003, MNRAS, 339, 289

Springel V., et al., 2005, Nature, 435, 629

Srianand R., Hussain T., Noterdaeme P., Petitjean P., Krühler T., Japelj J., Pâris I., Kashikawa N., 2016, MNRAS, 460, 634

Steinmetz M., Mueller E., 1994, A&A, 281, L97

Stengler-Larrea E. A., et al., 1995, ApJ, 444, 64

Sternberg A., Le Petit F., Roueff E., Le Bourlot J., 2014, ApJ, 790, 10

Stevens A. R. H., et al., 2019, MNRAS, 483, 5334

Stinson G., Seth A., Katz N., Wadsley J., Governato F., Quinn T., 2006, MNRAS, 373, 1074

Stinson G. S., Brook C., Macciò A. V., Wadsley J., Quinn T. R., Couchman H. M. P., 2013, MNRAS, 428, 129

Stone J. M., Gardiner T. A., Teuben P., Hawley J. F., Simon J. B., 2008, ApJS, 178, 137

Storrie-Lombardi L. J., Wolfe A. M., 2000, ApJ, 543, 552

Stott J. P., et al., 2014, MNRAS, 443, 2695

Tegmark M., Silk J., Rees M. J., Blanchard A., Abel T., Palla F., 1997, ApJ, 474, 1

Tescari E., Viel M., Tornatore L., Borgani S., 2009, MNRAS, 397, 411

Teyssier R., 2002, A&A, 385, 337

Teyssier R., Moore B., Martizzi D., Dubois Y., Mayer L., 2011, MNRAS, 414, 195

Thacker R. J., Couchman H. M. P., 2000, ApJ, 545, 728

Thacker R. J., Scannapieco E., Couchman H. M. P., 2006, ApJ, 653, 86

Theuns T., 2021, MNRAS, 500, 2741

Theuns T., Chalk A., Schaller M., Gonnet P., 2015, arXiv e-prints, p. arXiv:1508.00115

Thoul A. A., Weinberg D. H., 1996, ApJ, 465, 608

Tinker J., Kravtsov A. V., Klypin A., Abazajian K., Warren M., Yepes G., Gottlöber S., Holz D. E., 2008, ApJ, 688, 709

Tornatore L., Borgani S., Dolag K., Matteucci F., 2007, MNRAS, 382, 1050

Tytler D., 1982, Nature, 298, 427

Upton Sanderbeck P. R., D'Aloisio A., McQuinn M. J., 2016, MNRAS, 460, 1885

Veilleux S., Shopbell P. L., Miller S. T., 2001, The Astronomical Journal, 121, 198

Veilleux S., Maiolino R., Bolatto A. D., Aalto S., 2020, A&A Rev., 28, 2

Viel M., 2005, in Williams P., Shu C.-G., Menard B., eds, IAU Colloq. 199: Probing Galaxies through Quasar Absorption Lines. pp 255–260 (arXiv:astro-ph/0504645), doi:10.1017/S1743921305002681

Viel M., Haehnelt M. G., Springel V., 2004, MNRAS, 354, 684

Villaescusa-Navarro F., et al., 2018, ApJ, 866, 135

Vogelsberger M., et al., 2014, MNRAS, 444, 1518

Wadsley J. W., Stadel J., Quinn T., 2004, New A, 9, 137

Wadsley J. W., Veeravalli G., Couchman H. M. P., 2008, MNRAS, 387, 427

Wang Y., 2008, J. Cosmology Astropart. Phys., 2008, 021

Wang J., et al., 2014, MNRAS, 441, 2159

Wang W.-H., Kanekar N., Prochaska J. X., 2015, MNRAS, 448, 2832

Weinberger R., et al., 2017, MNRAS, 465, 3291

Wetzstein M., Nelson A. F., Naab T., Burkert A., 2009, ApJS, 184, 298

White S. D. M., Frenk C. S., 1991, ApJ, 379, 52

Whitworth A. P., Ward-Thompson D., 2001, ApJ, 547, 317

Wiersma R. P. C., Schaye J., Smith B. D., 2009a, MNRAS, 393, 99

Wiersma R. P. C., Schaye J., Theuns T., Dalla Vecchia C., Tornatore L., 2009b, MNRAS, 399, 574

Wilkins S. M., Feng Y., Di Matteo T., Croft R., Lovell C. C., Waters D., 2017, MNRAS, 469, 2517

Wolfe A. M., Prochaska J. X., 1998, ApJ, 494, L15

Wyithe J. S. B., Dijkstra M., 2011, MNRAS, 415, 3929

Xie L., De Lucia G., Hirschmann M., Fontanot F., Zoldan A., 2017, MNRAS, 469, 968

Xie L., De Lucia G., Hirschmann M., Fontanot F., 2020a, arXiv e-prints, p. arXiv:2003.12757

Xie L., De Lucia G., Hirschmann M., Fontanot F., 2020b, MNRAS, 498, 4327

Yajima H., Choi J.-H., Nagamine K., 2012, MNRAS, 427, 2889

Young P. J., Sargent W. L. W., Boksenberg A., Carswell R. F., Whelan J. A. J., 1979, ApJ, 229, 891

Young L. M., et al., 2011, MNRAS, 414, 940

Yung L. Y. A., Somerville R. S., Finkelstein S. L., Popping G., Davé R., 2019, MNRAS, 483, 2983

Zabl J., et al., 2020, MNRAS, 492, 4576

Zafar T., Péroux C., Popping A., Milliard B., Deharveng J. M., Frank S., 2013, A&A, 556, A141

Zafar T., et al., 2015, A&A, 584, A100

Zhu H., Avestruz C., Gnedin N. Y., 2020, ApJ, 899, 137

Zoldan A., De Lucia G., Xie L., Fontanot F., Hirschmann M., 2017, MNRAS, 465, 2236

Zwaan M. A., Meyer M. J., Staveley-Smith L., Webster R. L., 2005, MNRAS, 359, L30

van de Voort F., Schaye J., Booth C. M., Haas M. R., Dalla Vecchia C., 2011, MNRAS, 414, 2458

van de Voort F., Schaye J., Altay G., Theuns T., 2012, MNRAS, 421, 2809

van de Voort F., Springel V., Mandelker N., van den Bosch F. C., Pakmor R., 2019, MNRAS, 482, L85

van der Kruit P. C., Freeman K. C., 2011, ARA&A, 49, 301

van der Kruit P. C., Searle L., 1982, A&A, 110, 61